OUTLINE
Retinal repair

nature

THE INTERNATIONAL WEEKLY JOURNAL OF SCIENCE

ANCIENT HUMAN HYBRID

The first-generation offspring of a Neanderthal mother and a Denisovan father PAGE 113

ARTIFICIAL INTELLIGENCE

TWITTER AND THE STREETS

How tracking tweets could tackle gun violence

PAGE 20

PLANETARY SCIENCE

JUPITER'S DYNAMO

Unique nature of Jovian magnetic field revealed

PAGES 36 & 76

STRUCTURAL BIOLOGY

SURVIVAL STRATEGY

The protein complex that keeps malaria parasite alive

PAGES 41 & 70

NATURE.COM
6 September 2018
Vol. 561, No. 7721

THIS WEEK

EDITORIALS

SCHRÖDINGER Influence of physicist's "little book" continues to grow **p.6**

WORLD VIEW How lab manuals can offer help to young scientists **p.7**

MOTHER'S MILK The unlikely source of naked mole rat maternal instinct **p.9**

Proceed with caution

Proposed molecular testing of a person's age highlights difficult questions for scientists and society.

Eva Perón claimed she was younger for political reasons. Walt Disney pretended to be older to sign up for war. All sorts of people lie about their age for all sorts of reasons, and they've been doing so for a long time. Scandals over the past decade have forced authorities to act on false claims from footballers: the world governing body FIFA now routinely scans the wrist bones of players entering youth competitions to check that the athletes are truly young enough to compete.

Wrist scanning is one of several anatomical tests available that claim to be able to determine an individual's maturity. They are cumbersome and unreliable. More-accurate tests are on the way. But the implications are profound, and must be discussed by researchers, policymakers and the public.

One reason that scientists are trying to improve age testing is to address controversy in Europe over the age of refugees. United Nations rules say that those under 18 years old must receive particular protection and assistance. Some adults say they are younger than they are to claim these benefits, and tabloid anger over these rare cases has fuelled political and public intolerance.

If officials who assess age on the basis of physical attributes such as height, voice and facial features suspect that a refugee is concealing their true age, they can apply tests that assess the maturity of teeth or bones.

Experts are right to condemn these techniques as being unreliable and to complain they could deny vulnerable children the help afforded to them in national laws. Could a more-reliable and evidence-based test help? Such a test would not answer all the questions posed by Europe's refugee crisis; almost four million people have claimed asylum since 2014, sparking a rise of xenophobia in parts of society and creating difficult decisions about how best to help immigrants and who to help most. But accurate tests could, in theory, help to make the basis of these decisions objective and transparent.

As we report in a News story (see page 15), some scientists say that age could be determined more accurately using a molecular test called an epigenetic clock. It looks for distinctive chemical marks that are known to steadily accumulate on DNA. In theory, the test could be performed on a simple cheek swab. The researchers developing it are confident that this method could reliably predict age to within one or two years — much better than the currently used anatomical tests, which can estimate only to within three or four years. But they also highlight that the epigenetic clock currently performs poorly for many people. And no biological test will ever be able to say for sure if a person is 17 years, 11 months or 18 years old.

Such assays must be subjected to extensive, rigorous testing across different populations, and their limitations must be made clear. Furthermore, the ethical implications should be fully debated before the tests are used to determine the age of refugees, which potentially has life-changing consequences. Such tests must always come with full consent and privacy safeguards.

There are also important implications in other spheres. A more accurate way to determine age would be useful in forensics work, for example, to help build up a profile of a suspect from blood or semen. Success would depend on the sample being large enough for analysis. What's more, countries such as Germany currently prohibit such information being extracted from DNA tests.

Age fraud is a widespread problem in sport. In 2010, the discovery of such deception forced a Chinese gymnastics team to return the gold medals it had won at the Sydney Olympics a decade earlier. One competitor was declared to have been only 14 at the time of the competition, two years below the required age.

"Researchers should not develop, or make claims for, age-determining tests without extreme care."

Last week saw violent anti-immigrant demonstrations, this time in Chemnitz in eastern Germany. Even though the number of refugees arriving in Europe has plummeted — 49,000 had arrived by this July, compared with 1.3 million throughout 2016 — tensions remain high.

Given all this, researchers should not develop, or make claims for, age-determining tests without extreme care and wide discussion. That will take time: time for the rest of society to ponder how and whether such tests should be used. Age is not just a number. Much can be at stake. ■

Climate politics

Global warming tops the Australian agenda as climate debates depose a third prime minister.

Australia has two pressing environmental problems: climate change and finding a leader who can tackle it. Large swathes of the country are suffering the effects of a seven-year drought, the bush fire season has hit those parts two months early, and the destruction of the Great Barrier Reef grows more severe each year. Yet late last month, the country's attempts to make some modest changes to its energy policy to help reduce greenhouse-gas emissions blew up an internal storm in the ruling Liberal party that cost Prime Minister Malcolm Turnbull his job.

To lose one prime minister to political fights about climate-change policy is unfortunate. Two would be careless, but Turnbull is actually the third Australian premier to fall in this way in under a decade. What is going on? And what does this turmoil say about attempts to rein in damaging carbon emissions elsewhere?

All politics is local, and Australian climate politics more so than most. Although Australian scientists are world leaders in several areas

of climate science, including atmospheric monitoring of the Southern Hemisphere and understanding the causes of sea-level rise, the nation remains heavily reliant on coal for jobs and electricity. It mines more than half a billion tonnes of the stuff each year, and sells almost three-quarters of that abroad. The rest is burnt in Australian power stations, with electricity generation accounting for around one-third of the nation's greenhouse-gas emissions.

It's no coincidence that when Turnbull's political colleague (and then-treasurer) Scott Morrison wanted to criticize environmentalists last year, he brought a lump of coal to parliament and spoke about it in glowing terms. Last week — after Turnbull confirmed he was quitting politics — his son complained about the "undue level of influence" of the coal lobby. Morrison, who replaced Turnbull as prime minister, has yet to announce the fate of the disputed policy, the National Energy Guarantee, which would force emissions generators to show they are meeting annual standards. He has at least said that the country will not withdraw from the Paris climate agreement, a move being pushed by some government members.

He should stand firm. Although the Paris agreement is weak compared with the scale of what is needed, it represents a political triumph and one that places so few binding demands on nations that any withdrawal would be little more than crowd-pleasing theatrics. And most of the crowd won't be pleased: a June poll showed that 59% of Australians saw climate change as a pressing threat and one that needed action — the highest percentage in a decade.

A larger-scale survey last year of 38 countries showed a similar level of concern. But politicians in many of these places, even those fully behind the need for action on emissions, are also finding it difficult to follow through on pledges. Take Canada, where Justin Trudeau's government last month announced it was scaling back plans for a carbon tax. Last week, Nicolas Hulot, the French environment minister, resigned, claiming that governments around the world are not taking sufficient steps to tackle green issues such as climate change. And the reckless stance of US President Donald Trump continues to erode climate regulations and embolden climate sceptics. New Zealand, for one, still has ambitions for emissions-reducing laws, but many of the other promises the country made in Paris — including actual cuts to carbon emissions and boosts in foreign aid to help poorer countries adapt — are weakening under political pressure.

Many of those poorer countries are on the front line and will suffer heavily as the weather worsens. So will Australia. Droughts there are projected to increase in length and severity as a result of climate change. Heatwaves, floods and bush fires are also linked to global warming, and are predicted to become more common and more extreme. The country's island neighbours in the Pacific are likely to be inundated as sea levels rise. As a result, Australia, whose draconian refugee policy is a source of shame to many citizens, is likely to face an increase in climate refugees.

"Australia is likely to face an increase in climate refugees."

That these topics are now routinely debated amid mounting public concern about global warming is a victory of sorts for scientists, who must continue their efforts to make the case for action, and to research and speak out about the consequences. And although the current political drama in Australia paints a depressing picture, there is a glimmer of hope. A decade after the financial crash wrested away attention and momentum, climate change is once again at the top of the political agenda.

Things can change quickly in politics, and Australia has a chance to force that change. Already the opposition Labor party has promised a new emissions-reduction scheme. And next year, the country will again vote on its leader. For whoever wins that election, curbing climate change should be at the top of their to-do list — and they must be given the chance to hang around long enough to do so. ■

What is Life?

The lectures of physicist Erwin Schrödinger helped to change attitudes in biology.

In the winter of 1943, the physicist Erwin Schrödinger invited the Dublin public to hear him deliver a series of lectures he described as "difficult" and that "could not be termed popular". Some 400 people were undeterred and were among the first to hear Schrödinger offer his views on how physics could shed light on the puzzling ability of living organisms to maintain molecular order and organization in the face of what seemed to be the randomizing forces of nature.

Seventy-five years on, some of his ideas remain difficult — controversial even. But they are popular, and are once again drawing people to the Irish capital. Trinity College Dublin will this week host 'Schrödinger at 75 — The Future of Biology', at which a stellar cast of speakers will consider the future of disciplines ranging from ageing and plant science to infectious disease and consciousness.

Schrödinger's lectures were collected into what he called his "little book", *What Is Life?*, published in 1944 (see *Nature* **560**, 548–550; 2018). Some consider it one of the most influential scientific books of the twentieth century.

The book attracted scientists from other fields to the study of genetics and the molecular mechanisms of life, among them physicist Francis Crick and zoologist James Watson. But can the ideas in this slim volume really supply sufficient motivation for such a diverse programme?

Critics have rightly argued that the book was neither particularly original nor up to date. Schrödinger made the auspicious proposal that the genetic material is an "aperiodic crystal": a structure with a specific but not periodic arrangement of atoms, encoding information that somehow guides the development of the organism. That vision resonated with Crick and Watson as they contemplated the structure of DNA, but it wasn't wholly original. As to how the genetic machinery works, Schrödinger could only point out that it seems to suspend the second law of thermodynamics.

The impact of *What Is Life?* lies more in its spirit than its substance. Schrödinger presented the problem of life as a puzzle posed to no single discipline. And his timing was perfect: biology was already changing to a mechanistic and microscopic science. This cross-disciplinary relevance applies equally to the topics addressed at the Dublin meeting. The physical-sciences content of artificial intelligence and complex systems is obvious, but understanding of (say) cognitive neuroscience, learning and memory and infectious disease can also benefit from wide-ranging expertise: for example, from the study of network topologies, the thermodynamics of information, and ergodicity (how widely a dynamic system explores its available states).

Happily, chemistry is welcomed to this table too. That subject, after all, is what biologists relied on mid-century to probe and better understand DNA, enzymes and cell signalling. The subsequent emergence of molecular biology, due in large part to some of those inspired by *What Is Life?*, means that whether Nobel prizes get assigned to 'chemistry' or 'physiology or medicine' is now as arbitrary as whether Nobels in nuclear science in the early twentieth century were awarded in chemistry or physics.

What Is Life? made the case that profound questions about the natural world aren't owned by any academic discipline. Indeed, the Dublin meeting could have gone further by embracing Schrödinger's epilogue on determinism and free will, which invoked philosopher Immanuel Kant and Hinduism (and spoilt the book's chances of publication in devoutly Catholic Ireland). Some eyebrows were raised at this material, but Schrödinger's friend Albert Einstein would have seen nothing amiss in it. Philosophers, ethicists, poets and theologians also have a stake in the future of life. Perhaps they will be invited to the centenary. ■

The key to a happy lab life is in the manual

A well-crafted set of guidelines and advice can save time, reassure trainees and promote a positive lab culture, argues **Mariam Aly**.

A year and a half ago, as I was preparing to launch my own laboratory studying cognition at Columbia University in New York City, I kept returning to a particular concern: I would soon be responsible for the scientific advancement of trainees. How could I help them be the best scientists they could be, while also protecting their well-being?

I found the answer on Twitter. Two principal investigators in my field, Jonathan Peelle at Washington University in St Louis, Missouri, and Maureen Ritchey at Boston College in Massachusetts, shared their lab manuals. These laid out expectations for themselves and their trainees, as well as resources and tips to guide trainees through their time in the lab. I decided to follow in their footsteps by writing a lab manual to introduce my trainees to my philosophy for research and work–life balance. This required a great deal of time and thought, but it is something I would recommend to anyone leading a research group.

In the final few months of my postdoctoral studies, I thought about what had worked well and not so well for me as a trainee, and how to create best practices for my lab. Then I put into writing things that are usually transmitted informally. For example, that it doesn't matter to me whether trainees arrive at 9 a.m. or 1 p.m. or work from home, as long as they get their work done and honour their commitments. And I explicitly encouraged trainees to talk to me if they need to vent or feel they are foundering: academia can be stressful, and I want to help.

I addressed concerns that I imagined trainees would have: what if I make a mistake in my experiment? (It's OK, we all do; tell your collaborators right away so that you can start to discuss the next steps.) Do I have to work 80 hours a week to succeed? (No.) How do I ensure my results are reproducible? (Double-check your code, add explanatory comments, document every step of data analysis and use version control.) How do I participate in open science? (Publicly share stimuli, code and data when you submit a manuscript.)

I supplemented my lab manual (go.nature.com/2c1dxdt) with a wiki (go.nature.com/2pti9kj), a website of resources for lab members. This included everything from tools for learning the programming languages R and Python and how to do neuroimaging analyses, to tips on keeping up with the research literature (by using RSS feeds and Twitter) and where to find the best bagel in Manhattan (a ten-minute walk from the lab). My goal was that any newly accepted lab member could read the manual and wiki and then strut into the lab knowing what to expect.

I try hard to keep myself accountable for what I have written. For instance, I promised weekly meetings with each trainee, and I stick to that, although it's a challenge with teaching obligations and travel. I hope that the consistency between my actions and my words helps lab members to understand that I meant what I wrote, even if they have yet to experience everything I promised.

I ask every trainee to read the lab manual. I make a point of referencing it and the wiki, along with repeated, not-so-subtle examples of their utility. My lab members now contribute to the wiki without prompting. When I checked in with them to see whether these resources were useful, the answer was a resounding 'yes'. Their actions also suggest that they internalized what they read. Some share struggles with me, ask for advice and take days off for mental health — as I hoped they would, and as I wish I had done when I was a trainee.

Here's another example: my lab manual states that trainees are entitled to read my grants, and my lab members have requested to see them. That's something I never asked my previous advisers; I worried it would be presumptuous. I realize now that my thinking was almost certainly wrong, but my own uneasy feelings as a trainee just drive home how important it is to put into writing that something is OK — otherwise, trainees might assume it is not. That goes double for the areas that trainees are most sensitive about: I've written down in black and white that it is OK to make mistakes and to maintain a work–life balance.

MY TRAINEES DO NOT **HAVE TO STRUGGLE** TO FIND ANSWERS TO **COMMONLY ASKED** QUESTIONS.

Putting together a lab manual and wiki takes time, but there are several examples to use for inspiration. My lab manual and wiki are publicly available for anyone to use as starting points. Once the wiki has been written, the entire lab can help to maintain it; if everyone pitches in, any particular update will often take only a few minutes.

The initial effort of writing a manual saves enormous amounts of time in the long run. I no longer have to repeatedly search my e-mails or the Internet to find the answer to a problem I previously solved but have forgotten. Likewise, my trainees do not have to struggle to find answers to commonly asked questions (for example, 'how do I get after-hours access to the building?'). More importantly, having a lab manual requires you to be explicit and transparent about your expectations and what you promise to do for your lab — every trainee reads the same expectations in the same words, putting everyone on equal footing.

A year after writing the lab manual, I re-read and revised it. That process reminded me of all that was at stake: all that I promised my trainees and all that I needed to do to ensure a healthy, happy and safe lab environment. It also led me to reflect on how pleased I am with my lab. My trainees are hard-working, sociable and supportive of one another. I love walking in and seeing them working together on a problem, or laughing and dancing when they've solved one. I might have written the lab manual, but my trainees brought it to life. ■

Mariam Aly *is an assistant professor in the Department of Psychology at Columbia University in New York City.*
e-mail: ma3631@columbia.edu

SEVEN DAYS *The news in brief*

HEALTH

STDs on the rise

The incidence of several sexually transmitted diseases is rising steadily in the United States, the US Centers for Disease Control and Prevention (CDC) said on 28 August. Nearly 2.3 million cases of gonorrhoea, syphilis and chlamydia were diagnosed last year, up from 2.1 million in 2016. The number of people diagnosed with gonorrhoea jumped by 67% between 2013 and 2017; syphilis diagnoses rose by 76% and chlamydia by 22%. Without prompt treatment, these infections can cause infertility and stillbirth, and increase the risk that a person will contract HIV. The CDC also says that the bacteria that cause gonorrhoea are becoming more resistant to antibiotics.

ENERGY

Carbon-free energy

The California state legislature passed a bill on 29 August that would require the state to generate 100% of its electricity from carbon-free sources by 2045. It would also increase the state's existing carbon-free-electricity mandate for 2030 from 50% to 60%. Governor Jerry Brown must now sign the bill before it can become law. Once the regulation is enacted, California will become the second US state to establish such a policy. Hawaii was the first, enacting a similar mandate in 2015. Massachusetts, New Jersey, New York and Washington DC are also considering carbon-free-electricity mandates.

EVENTS

Bullying probe

An investigation into allegations of bullying is under way at the prestigious Wellcome Sanger Institute in Hinxton, UK. According to *The Guardian*, which reported on the investigation on 29 August, ten former and current staff members have accused leaders at the institute of bullying, mistreatment of staff and gender discrimination. The complaints include allegations against the institute's director, geneticist Michael Stratton. The investigation will "ensure full and proper exploration of these allegations", an institute spokesperson said. The institute is owned by the Wellcome Trust, which unveiled a landmark anti-bullying policy in May. The trust said that it is aware of the investigation at Sanger, and that it would "await the outcome of that process before commenting further".

RICARDO MORAES/REUTERS

Brazil's national museum gutted by fire

A huge fire devastated Brazil's National Museum in Rio de Janeiro on the evening of 2 September. Many of the archaeological finds and historical memorabilia — some 20 million items in all — are now feared to have been destroyed; museum officials told local media that as little as 10% of the collection might have survived. The 200-year-old building housed several landmark collections, including Egyptian and Greco-Roman artefacts and the oldest human-skull fossil found in the Western Hemisphere. It's not yet clear what caused the fire, but the lack of a sprinkler system, the dilapidated state of the building and the failure of the two fire hydrants closest to the museum have all been blamed for the extent of the damage. Brazil's education minister, Rossieli Soares, told reporters on 3 September that the federal government would spend an initial 15 million reais (US$3.6 million) to restore the structure and rebuild its collection, and would seek international help.

Research-ship clash

Japan's foreign ministry lodged a complaint with the South Korean government on 28 August, after a Korean research ship was spotted near a group of small islands that both countries claim as their own. The islets, known as Dokdo in Korea and Takeshima in Japan, are controlled by South Korea, but sit in a joint fishing zone where the nations' claims also overlap. Japan asserts that the ship, *Tamgu 20*, was seen in Japan's exclusive economic zone southwest of the islets, where foreign research vessels require permission to operate. According to Japan's coast guard, the ship — managed by South Korea's National Institute of Fisheries Science (NIFS) in Busan — dropped a black object into the sea and left the zone before noon. The NIFS has confirmed that the ship was conducting marine research on that day, but gave no details of its location or activities.

Salk lawsuit claims

A California court has thrown out a retaliation claim in a gender-discrimination lawsuit

against the Salk Institute for Biological Studies. On 30 August, a judge dismissed molecular biologist Beverly Emerson's claim that the institute in La Jolla, California, let her contract expire in December 2017 because of the suit she had filed that July. The court also ruled that a key piece of evidence for the claim — an e-mail from Salk's former president, Elizabeth Blackburn, suggesting that litigation might hurt Emerson's career — is confidential material that should not go before jurors. Emerson alleges that systemic bias at the institute limited her pay and professional advancement and blocked her from resources such as research funding. The gender-discrimination trial is scheduled for 7 December.

PEOPLE

French minister

François de Rugy, speaker of the French National Assembly, was appointed France's environment minister on 4 September. De Rugy is a former green-party politician who switched to President Emmanuel Macron's party, La République En Marche, in last year's legislative elections. His predecessor, Nicolas Hulot, resigned his post on 28 August during a dramatic live

ALEXANDROS MICHAILIDIS/SHUTTERSTOCK

interview on the radio station France Inter. Hulot (**pictured**), a popular environmental activist who became Macron's environment minister last May, had declined previous offers to join government, saying that he had more influence as an activist. But a year after taking the job, Hulot expressed frustration with the slow pace of progress in politics, and said that short-term demands on government detracted from its ability to tackle long-term issues such as climate change and the decline of biodiversity and natural resources.

SPACE

Mars-rover vigil

NASA is waiting for a massive dust storm on Mars to ebb before attempting to waken the sleeping Opportunity rover, the agency said on 30 August. The 14-year-old rover went silent on 10 June, after the storm obscured the sunlight that the spacecraft needs to survive. When the dust abates, NASA will listen for signals from the rover, and send it messages, for at least 45 days. The agency will continue to listen in a more passive mode until at least January 2019. By then, the Martian summer will have given way to autumn at Opportunity's landing site. That change in seasons could kick up windstorms and clear any dust coating the rover's solar panels. The recent Martian dust storm is one of the most extensive ever seen on the red planet.

RESEARCH

Higgs decay

The ATLAS and CMS experiments at the Large Hadron Collider (LHC) have observed a previously undetected way in which the Higgs boson can decay — into a particle called the bottom quark, and its antiparticle. The experiments, based at CERN, Europe's particle-physics laboratory outside Geneva, Switzerland, discovered the Higgs, a key part of the mechanism that gives other particles their masses, in 2012. LHC researchers have accumulated evidence of the particle decaying into a variety of products, following theoretical predictions, including decaying into two photons. In June, researchers revealed that they had also seen the Higgs interact with the top quark. The bottom-quark decay, announced on 28 August, is expected by theory, but the signal had been difficult to single out from the many other processes that can also produce those particles.

Retraction report

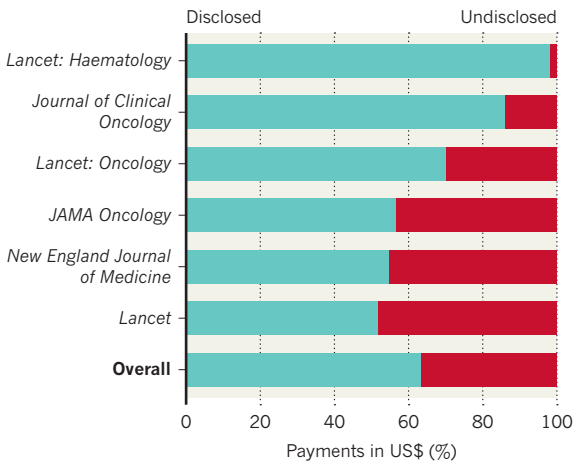
A Chinese university has concluded that the authors of a controversial gene-editing paper that was later retracted did not intend to deceive the scientific community. The paper detailed how an enzyme called NgAgo could edit genomes with similar accuracy to the CRISPR–Cas9 gene-editing system (F. Gao *et al. Nature Biotechnol.* **34**, 768–773; 2016). But the paper's main finding was within months challenged by scientists who failed to reproduce the results. On 2 August last year, the authors agreed to retract the paper. Last week, Hebei University of Science and Technology in Shijiazhuang announced that its investigation found no basis for thinking that the original experiments should be republished. (*Nature*'s news team is editorially independent of *Nature Biotechnology*.)

TREND WATCH

An analysis of cancer-drug trials published in leading journals finds that about one-third of US authors failed to fully disclose payments from trial sponsors. Researchers led by Cole Wayant, a meta-researcher at Oklahoma State University in Tulsa, searched for cancer drugs approved by the US Food and Drug Administration between January 2016 and August 2017. They then looked for related clinical trials and identified the key publication for each trial. For any US physician-oncologist authors named on a study, Wayant's team looked at whether the authors had received funding from a trial sponsor during the period of the trial. Under the US Affordable Care Act, such payments must be declared by drug companies in the publicly available Open Payments database. The researchers cross-referenced these payments with conflict-of-interest declarations made by the authors in the journal articles, and added up the total amounts disclosed and undisclosed for each article. Of the 344 oncologist authors for whom the researchers had data, 110 (32%) had not fully disclosed payments from the sponsor. Together, these authors received about US$217 million in payments.

CANCER-DRUG PAYMENTS

One-third of 344 oncologist researchers involved in drug trials published in leading oncology journals failed to fully disclose payments from companies that sponsored the studies.



SOURCE: C. WAYANT *ET AL. JAMA ONCOL.* HTTPS://DOI.ORG/10.1001/JAMAONCOL.2018.3738 (2018).

NEWS IN FOCUS

SIVARAM V/REUTERS

Indian soldiers rescue residents from a flooded area in the southern state of Kerala.

SUSTAINABILITY

Kerala floods made worse by mining and dams

Scientists say development boom in the Western Ghats mountains contributed to the disaster.

BY T.V. PADMA

Torrential rains pounded southwest India in August, triggering devastating floods in the state of Kerala that have so far killed at least 483 people and forced hundreds of thousands from their homes. The monsoon rains have been heavier than usual, but scientists say that outdated dam-management systems and increasing mining and development in the Western Ghats mountain range — a biodiversity hotspot that ecologists are trying to conserve — have exacerbated the disaster.

Kerala received 758.6 millimetres of rain between 1 and 19 August — 2.6 times the average for that time of year. The unusually heavy downpours caused rivers to overflow. Many of the fatalities were the result of landslides in rural areas, triggered by the massive downpours. Authorities say the floods are the state's most damaging in 100 years.

A contributing factor is that after the heavy rain, authorities began to release water from several of the state's 44 dams, where reservoirs were close to overflowing. The neighbouring state of Tamil Nadu also purged water from its over-filled Mullaperiyar dam, which wreaked yet more havoc downstream in Kerala.

Scientists say state governments often allow reservoirs to fill completely early in the monsoon season, and do not release water slowly at regular intervals to prevent overfilling later in the season. "India's reservoir management is ▶

unscientific," says meteorologist Madhavan Nair Rajeevan, secretary of India's ministry of Earth sciences, which oversees the country's meteorological institutes. Computer models and meteorological forecasts are used in Europe and the United States to predict the rate at which water flows into reservoirs and how much water needs to be stored — but few authorities in India use such systems, says Rajeevan. He suggests that prediction systems should be introduced across India.

An increase in development over the past two decades in the Western Ghats — a large mountain range that runs parallel to India's west coast and across several states — might also have exacerbated the flooding, say several Indian scientists. "The land and soil disturbances have triggered landslides and blocked streams, contributing to the floods," says Madhav Gadgil, an ecologist at Goa University in Taleigão.

In 2011, Gadgil headed a committee that investigated environmental damage from unsustainable development and illegal mining in the Western Ghats. The committee recommended that the entire mountain range be declared "ecologically sensitive" — it contains 30% of India's plant, fish, bird and mammalian species — and that mining and the construction of dams and coal-fired power plants be banned to conserve biodiversity.

But the government ignored the report's recommendations. Instead, in 2013, it accepted the advice of another committee, which suggested that only 37% of the Western Ghats be made off-limits to mining and construction.

Gadgil says that the state governments have continued to approve infrastructure projects across the Ghats, including dams, power plants and buildings, many without reliable environmental-impact assessment reports. "There has been a proliferation of building and road construction," he says. He adds that there's also been an increase in illegal mining.

Jason von Meding, who studies disaster risk-reduction at the University of Newcastle in Australia, says the government should explain why it rejected the Gadgil-committee report, which emphasized the need to curb development excesses and focus on conservation. "Uncontrolled mining, dam construction, deforestation and poorly planned construction have multiplied the risk of flooding and landslides in recent years," he says.

Earth scientist Rajiv Sinha at the Indian Institute of Technology Kanpur also points to an increase in the numbers of canals and bridges, which can reduce the width of rivers, leading to a build-up of sediment that slows water flow. After a sudden downpour, there is not enough space for the water, so it floods the surrounding area, "leading to disasters like the one we are witnessing in Kerala", says Sinha. India's poor infrastructure planning will be exacerbated by its vulnerability to extreme rainfall events, which are projected to happen more frequently as a result of global warming, he says. ■

PUBLIC HEALTH

Experimental Ebola drugs face tough test

Researchers are devising a clinical-trial protocol to test three medicines in Africa's latest outbreak.

BY AMY MAXMEN

Health workers fighting the ongoing Ebola outbreak in the Democratic Republic of the Congo (DRC) have given nearly 20 people experimental drugs to treat the virus since mid-August. But because the drugs have been dispensed on a case-by-case, 'compassionate use' basis, it is hard to know whether any are effective. Now, desperate to determine which therapy works best, researchers from the DRC and US governments, the World Health Organization (WHO) and other groups are meeting to plan a clinical trial that will compare multiple drugs as the outbreak continues.

For ethical reasons, the trial scientists do not intend to give any study participants a placebo. Instead, they hope to compare the two experimental medicines already in use to ZMapp, an antibody therapy that showed promise three years ago during a major Ebola epidemic in West Africa (The PREVAIL II Writing Group. *N. Engl. J. Med.* **375**, 1448–1456; 2016). Patients in the coming trial would receive one of these three drugs at random. The study design draws on a flexible clinical-trial framework that the WHO expects to unveil this week. The framework is intended for use in multiple Ebola outbreaks, to produce data that can be pooled over time.

"Every trial has its own challenges."

The scientists working on the DRC trial hope to launch the effort in the coming weeks. "A clinical trial will give us the scientific evidence we need," says Jean-Jacques Muyembe-Tamfum, director-general of the National Institute for Biomedical Research in Kinshasa, which will lead the study.

But planning for the trial is complicated by the realities of working in the DRC's North Kivu and Ituri provinces, where fighting has killed more than 5 million people over the past two decades. Instability in the region could prevent clinicians from carrying out their work. "Armed groups can do what they want," Muyembe-Tamfum says.

The current outbreak began on 1 August, and has grown to include 115 confirmed and probable cases of Ebola — including 77 people who have died, the DRC health ministry said on 28 August. Public-health workers have vaccinated 4,645 people, and doctors have given 3 people the antiviral drug remdesivir, made by Gilead Sciences of Foster City, California. Another 13 patients have received mAb114, an experimental treatment derived from antibodies found in the blood of a person who contracted Ebola in 1995 and survived.

That swift response is a major shift from the handling of the Ebola epidemic that struck West Africa in 2014. Experimental drugs were not used widely in West Africa then because there was no proof of their safety or efficacy — clinical trials did not begin until the outbreak was near its end. That delay helped to drive the death rate among Africans infected with Ebola to 63%. But several Westerners infected with Ebola received the nascent therapies at top hospitals; the fatality rate for this group of patients was just 18%. This disparity eventually prompted the WHO to develop guidelines aimed at ensuring wider access to experimental treatments during future Ebola outbreaks.

But the only way to determine how well a drug works is through a randomized, controlled clinical trial. Thus far, researchers have not managed to complete a trial of any experimental Ebola drug, because outbreaks of the disease have ended before enough patients enrolled. So the WHO has been working with international experts to create a basic trial design that can be adapted as data accumulate and logistical challenges change.

Muyembe-Tamfum says that the trial being planned now will use that framework, and is likely to test mAb114, remdesivir and ZMapp, made by Mapp Biopharmaceutical in San Diego, California.

Stationing a large number of medical professionals in an Ebola unit now is particularly fraught because there are more than 100 militias roving the eastern DRC. If armed groups show up at a treatment centre, workers might flee rather than risk their lives — and any trial could stop. But Ana Maria Henao Restrepo, who helps to lead the WHO's Ebola Research and Development team, is unfazed. "Every trial has its own challenges," she says. "That's why we are coming out with an approach that's flexible." ■

NEMANJA PANCIC/SIPA/SHUTTERSTOCK

Around 4 million refugees have arrived in Europe since 2014, many of whom have no identity documents.

POLITICS

DNA clock may aid refugee age check

European forensic scientists want to find out whether epigenetics can help determine which refugees are under 18.

BY ALISON ABBOTT

When authorities in Hildesheim, Germany, didn't believe an asylum seeker who claimed to be under 18 years old — and thus eligible for privileged treatment — police turned to a blood test sold by Zymo Research in Irvine, California.

The test uses chemical modifications to DNA that accrue over a lifetime, called an 'epigenetic clock', to determine a person's age. Scientists — aware of its potential benefits but also of its current lack of precision — sounded an alarm.

In a paper published[1] in May, a team led by forensic-medicine specialist Stefanie Ritz-Timme of the University of Dusseldorf in Germany said that these tests were not ready for use in sensitive forensic evaluations.

But now, in the charged political atmosphere that has accompanied the arrival of millions of refugees in Europe, researchers are joining forces to improve epigenetic-clock-based tests — with a focus on whether they might be used to help determine the age of refugees whose claims to be under 18 are disputed.

"The race is now on to develop a more accurate clock that would be more predictive than the anatomical tests — and also more practical," says cell biologist Wolfgang Wagner at the University of Aachen, Germany.

The development of methods that could feed into decisions about who is granted asylum and how refugees are treated are likely to elicit criticism, says Denise Syndercombe-Court, a forensic geneticist at King's College London. She says that some scientists, herself included, are wary of these efforts.

But Niels Morling, a forensic geneticist at the University of Copenhagen who is running a national epigenetic-clock programme, defends the research. Given that the law treats those under 18 very differently from adults, he says, "you have a duty to make sure that it can be implemented fairly".

Philosopher Thomas Pogge, who specializes in global justice at Yale University in New Haven, Connecticut, says that, to keep rising anti-immigration sentiment in check, it is important for authorities to show that they can detect any refugees who pretend to be younger than they are.

Since 2014, around 4 million refugees have arrived in Europe, many without identity documents. Minor status usually leads to better care, an increased chance of being granted asylum and a higher chance of gaining permission to be joined by family members.

Authorities say that some unaccompanied refugees claim to be younger than they are. But the anatomical tests that are currently used in some countries to assess age have an error range of up to 3–4 years and rely on X-rays and magnetic resonance imaging. In Norway, refugees have sued authorities for being forced to submit to these medical tests.

The publication of the first reasonably accurate epigenetic clock in 2013 presented a simpler way of verifying age, because the test could be done using blood samples[2]. Developed by biostatistician Steve Horvath at the University of California, Los Angeles, this clock measured an epigenetic mark called methylation at 353 DNA sites across the genome.

In July, Horvath and his team published a new clock that measures epigenetic marks at 391 DNA sites[3]. It was particularly accurate in buccal cells scraped from the inside of the cheek, which are easier to collect than blood samples. Testing these cells from 53 people aged between 3.5 and 18 years, he found a median error of just 1.03 years. However, there were many outliers — people whose age could not be accurately predicted — and the most extreme result was out by 5 years and 8 months. Horvath expects that epigenetic clocks, once refined, will help refugees by corroborating their age claims. "At the same time, the tests may help identify individuals who break the law," he adds.

In the Hildesheim case, Zymo, which bought an exclusive licence to Horvath's test in 2016, compared the refugee's sample with those of five others who had similar ethnic backgrounds and whose ages were known. Keith Booher, project manager for Zymo's epigenetic services, told *Nature* that the test determined the most likely age of the person to be between 26 and 29 years old. The Hildesheim authorities have declined to comment on the case.

Forensic scientists around Europe are now working to make an epigenetic clock that would be more accurate and less expensive than what is available — and that is applicable to the easily collected buccal cells. Studies are also under way to gauge how the diverse ethnic backgrounds of Europe's refugees might influence the epigenetic clock. Another challenge to using the clock is dealing with the people who are statistical outliers. Wagner thinks that some of these people, for whom the method simply won't work, could be identified using the right combinations of epigenetic markers, allowing scientists to discount the test's results.

Even if the accuracy of the epigenetic clock cannot be improved to rival anatomical tests, it would still be attractive, says Morling. "Every time you add a new method, you improve overall precision of age estimation." ■

1. Ritz-Timme, S., Schneider, P. M., Mahlke, N. S., Koop, B. E. & Eickhoff, S. B. *Rechtsmedizin* **28**, 202–207 (2018).
2. Horvath, S. *Genome Biol.* **14**, R115 (2013).
3. Horvath, S. *Aging* **10**, 1758–1775 (2018).

RESEARCH CAPACITY

World Bank invests in Africa

Grassroots science initiative receives US$280 million.

BY LINDA NORDLING

A World Bank scheme aimed at building research capacity in Africa has announced a third, and probably final, investment worth at least US$280 million. The initiative loans money to African governments, and has set up 46 education and research centres in 17 African countries — but some worry what will happen once the bank's money runs out.

"I see a big challenge when the funding ends," says Patrick Ogwang, who leads a traditional-medicine research centre funded by the initiative, at Mbarara University in Uganda. He is eyeing industry partnerships as a source of future cash, but says that competition is fierce.

The World Bank launched the African Centres of Excellence (ACE) initiative in 2014 with $165 million in loans; it created 22 centres in West and central African nations. Two years later, the bank approved $148 million to create 24 hubs in eastern and southern African countries. The third round, announced on 31 August, pushes the bank's total investment past $500 million. It again targets West and central Africa, and French development agency the AFD may add another $50 million.

The centres focus on local research challenges such as plant breeding and infectious diseases, and have created jobs for hundreds of scientists and trained thousands of graduate students. Centres are eventually expected to sustain themselves with funding from governments, charities and industry. It's important that the centres move towards this, says World Bank economist Andreas Blom, who leads the programme, but the third round of loans will offer "weaning off" funding for existing centres in West and central Africa, and pay for new ones.

Critics of the scheme say that it has allowed governments to delay making substantive national investments in research. Governments have 40 years to repay the money at low or zero interest. "Many African governments, with short political life spans, are not really concerned about who will pay and how," says John Mugabe, an expert on science policy in Africa at the University of Pretoria in South Africa.

Representatives from the Ghanaian and Nigerian governments told *Nature* that the ACE loans complement their plans for national funding. In Ghana, higher-quality research proposals will reach the country's national research fund thanks to the scheme, says Mohammed Salifu, executive secretary of Ghana's National Council for Tertiary Education. ■

Firefighters battle a blaze near Redding, California, in July.

MARK RALSTON/AFP/GETTY

ECOLOGY

Huge wildfires defy explanation

Researchers scramble to improve wildfire models as blazes become larger and less predictable.

BY JEFF TOLLEFSON

In California, where the state's largest wildfire on record continues to burn, fires are getting bigger and less predictable — so much so that scientists are struggling to model them. Now, two research projects under way in the state are aiming to revamp the models that scientists, first responders and policymakers use to understand these dangerous and costly disasters.

One, slated to wrap up in the next few months, looks at how specific environmental factors such as extreme winds affect fires. The other, officially launched on 30 August, focuses on how wildfires will change in the coming decades as the climate warms.

"Something is definitely different, and it raises questions about how much we really know," says Max Moritz, a fire scientist at the University of California, Santa Barbara.

The efforts come against a backdrop of abnormal fire seasons around the world. The giant California fire has torched about 166,000 hectares since late July, and continues to burn in the northern part of the state. British Columbia in Canada is now experiencing its worst fire season on record (see 'Scorched earth'). And in late July, after weeks of intense heat and some of the lowest rainfall totals since the late nineteenth century, officials in Sweden were battling roughly 50 wildfires across the country.

"We need to refocus some of our research efforts on characterizing the kinds of fire behaviour that cause us the most grief."

Researchers have been at a loss to explain a flurry of unusual fire behaviour in California in recent years: wildfires that burn hot throughout the night instead of settling down, as many used to; blazes that race down hillsides faster than before; and fires that torch suburban neighbourhoods that were once considered safe from such

events. And, in July, a tornado with unprecedented wind speeds of 230 kilometres per hour span up inside a fire near Redding, California.

The problem, Moritz says, is that most of the fire models in use today are based on data from the past two or three decades. But it seems that fire behaviour might be shifting in response to climate faster than anybody expected, and that makes it increasingly problematic to extrapolate from past trends, he adds.

BAD BEHAVIOUR

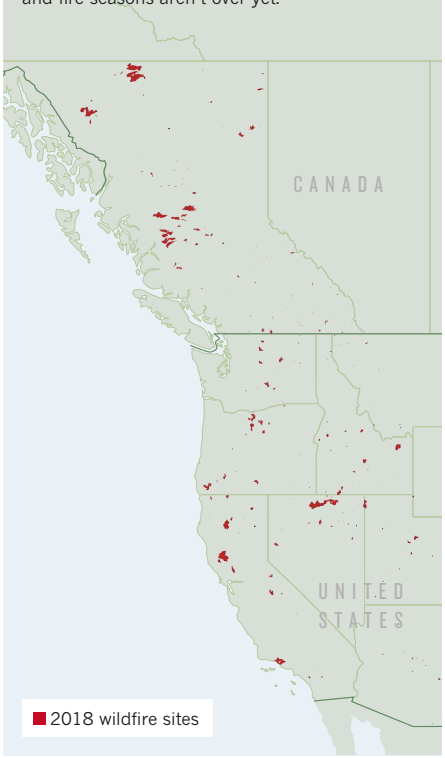
"More frequent, extreme fire behaviour is actually sort of expected, but just saying that it's going to happen isn't enough," says Dave Sapsis, who specializes in fire modelling and behaviour at the California Department of Forestry and Fire Protection (CAL FIRE), based in Sacramento. "We need to refocus some of our research efforts on characterizing the kinds of fire behaviour that cause us the most grief."

As part of one of the projects, Sapsis is updating the model that CAL FIRE uses to map fire hazards across the state. In use since 2007, the model incorporates information about environmental conditions such as topography, fire history and the type of burnable vegetation in an area. But it doesn't capture how extreme winds can move through a local landscape. Those winds are the key to understanding urban conflagrations, Sapsis says.

Within the next few months, he hopes to complete work on a detailed record of wind speed and direction across the entire state over the past 15 years. Those wind maps should help scientists to study recent fires and, ultimately, boost CAL FIRE's ability to predict the risk of extreme fires in any given locality, Sapsis says.

SCORCHED EARTH

Blazes in the United States and Canada have burnt nearly 4 million hectares this year — and fire seasons aren't over yet.



SOURCE: WRI/USGS/BC WILDFIRE SERVICE

Climate scientists expect those risks to increase in the coming decades. California's Fourth Climate Change Assessment, released on 27 August, projects that the area of land consumed by wildfires in the state each year could increase by 77% by 2100 if global greenhouse-gas emissions continue to rise. On average, more than 286,000 hectares have burnt each year over the past two decades.

FUTURE ON FIRE

The second project, a US$4-million study that includes Moritz and other scientists at multiple University of California campuses, will explore the future of fire, ecosystems and climate in California. Much of the existing research has focused on extrapolating from past trends. But this study is aiming to create a more realistic picture of how wildfires and ecosystems will evolve by integrating detailed models of fire behaviour, vegetation and climate across the entire state.

This should allow scientists to analyse how more-extreme and variable weather will affect wildfires and how ecosystems will respond to them, says Alex Hall, a climate scientist at the University of California, Los Angeles, and the project's principle investigator.

A lot of work has focused on tracking average fire trends, Sapsis says. But scientists need to improve their understanding of the extreme blazes, as well as how fire patterns could shift in the future, he adds. This will help government agencies and communities make better choices when it comes to managing ecosystems and human developments in fire-prone areas. ■

PUBLISHING

Radical plan to end paywalls

Top European research funders announce 'Plan S' to make all scientific works free to read.

BY HOLLY ELSE

Research funders from France, the United Kingdom, the Netherlands and eight other European nations have unveiled a radical open-access initiative that could change the face of science publishing in just two years — and which has instantly provoked protest from publishers.

The 11 agencies, which together spend €7.6 billion (US$8.8 billion) in research grants annually, say they will mandate that, from 2020, the scientists they fund make resulting papers free to read immediately on publication. The papers would have a liberal publishing licence that would allow anyone else to download, translate or otherwise reuse the work. "No science should be locked behind paywalls!" says a preamble document that accompanies the pledge, called Plan S, released on 4 September.

"It is a very powerful declaration. It will be contentious and stir up strong feelings," says Stephen Curry, a structural biologist and open-access advocate at Imperial College London. The policy marks a "significant shift" in the open-access movement, which has seen slow progress in its bid to make scientific literature freely available online.

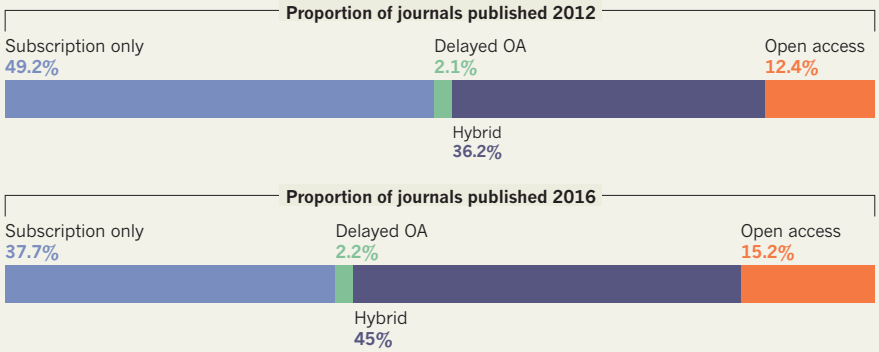
As written, Plan S would bar researchers from publishing in 85% of journals, including influential titles such as *Nature* and *Science*. According to a 2017 analysis, only around 15% of journals publish work immediately as open access (see 'Publishing models') — financed by charging per-article fees to authors or their funders, by negotiating general open-publishing contracts with funders, or through other means.

More than one-third of journals still publish papers behind a paywall, and typically permit online release of free-to-read versions only after a delay of at least six months. And just less than half have adopted a 'hybrid' model of publishing, whereby they make papers immediately free to read for a fee if an author wishes, but keep most studies behind paywalls. Under Plan S, however, scientists wouldn't be allowed to publish in these hybrid journals, except during a short transition period. The plan also states that funders will cap the amount they are willing to pay for open-access publishing fees, but doesn't lay out what charge would be too much.

The initiative is spearheaded by Robert-Jan Smits, the European Commission's special ▶

PUBLISHING MODELS

Worldwide, the proportion of subscription-only journals* shrank between 2012 and 2016, giving way to more open-access (OA) and hybrid journals.



*From Scopus database. Hybrid journals are subscription titles that allow authors to make individual papers open for a fee. Percentages do not add up to 100% because of rounding.

SOURCE: UNIVERSITIES UK

envoy on open access, and was launched by the advocacy group Science Europe (the 'S' in Plan S can stand for 'science, speed, solution, shock', Smits says). National agencies in Austria, Ireland, Luxembourg, Norway, Poland and Slovenia have also signed, as have funders in Sweden and Italy.

Smits says he took inspiration from the open-access policy of the Bill & Melinda Gates Foundation, the global health charity based in Seattle, Washington, which also demands immediate open-access publishing. Because Plan S forbids hybrid publishing — and because it involves multiple funders — its impacts could be more far-reaching than the Gates policy.

Despite Smits' role, the European Commission hasn't itself signed the plan. But Smits says that he expects the requirements to be integrated into the terms of future research grants from the commission. He also expects more funding agencies to join, and says he will discuss the plan in the United States next month.

Asked for comments on the plan, publishers said they had serious concerns — particularly around the banning of hybrid journals. A spokesperson for the International Association of Scientific, Technical and Medical Publishers (STM) in Oxford, UK, which represents 145 publishers, told *Nature*'s news team that it welcomed funders' efforts to expand access to peer-reviewed scientific works, but that some sections of Plan S require careful consideration to avoid "any unintended limitations on academic freedoms". In particular, the spokesperson said, banning hybrid journals — which have broadened the availability of open-access articles — could "severely slow down the transition".

The publisher Elsevier said it supported the STM's comments. A spokesperson for Springer Nature said: "We urge research funding agencies to align rather than act in small groups in ways that are incompatible with each other." Removing publishing options from researchers "fails to take this into account and potentially undermines the whole research publishing system", the statement added. (*Nature*'s news team is editorially independent of its publisher.)

Curry cautions that shifting from a subscription to an open-access business model around the world could also bring a new challenge — how scientists in poorer nations will be able to afford to publish open-access work. "That has to be part of the conversation," he says. ■

A WINDOW ONTO GUN VIOLENCE

Gang-related shootings plague many US cities, and researchers are trying to tackle the problem using artificial intelligence.

BY ROD MCCULLOM

In the middle of the day on 11 April 2014, a hooded gunman ambushed Gakirah Barnes on the streets of Chicago's South Side. A volley of bullets struck her in the chest, jaw and neck. The 17-year-old died in a hospital bed two hours later.

To many, her death was just another grim statistic from a city that has been struggling with gun violence. Last year, around 3,500 people were shot in Chicago, Illinois, of which 246 were aged 16 or younger; 38 of those children never celebrated another birthday.

But Barnes's death was unusual for several reasons. She was a young woman in an epidemic of violence that largely affects black men. She also had an Internet following. Barnes had a reputation as a 'hitta' — or killer — with rumours of at least two dead bodies to her credit. Although never charged with murder, she embraced the persona, posing in photos and videos with guns in her hands and making threats against rival gangs on Twitter. In a morbid modern irony, it's likely that she revealed her location in real time to her killer through a tweet. Police have yet to charge anyone in connection with her murder.

Desmond Upton Patton was sitting in his office at the University of Michigan in Ann Arbor when he first saw the headlines about Barnes. The social worker had been studying 'Internet banging', or 'cyberbanging', the use of social media by gang-involved youths to challenge, taunt or threaten rivals[1]. The online disputes can often spill out into the streets as physical violence.

Patton took a deep dive into Barnes's archived Twitter timeline and discovered a treasure trove of social-media data — random thoughts as well as boasts, threats and violent imagery. But what surprised him most, he says, was the grief. "My pain ain't never been told," Barnes wrote after a friend was killed just weeks before her own death.

What emerged from her timeline was a picture of a teenage girl who lived in a community steeped in violence, who was deeply hurt by it and who wanted revenge. Now at the Columbia School of Social Work in New York City, Patton thinks that social-media histories such as that of Barnes can offer ways to identify young people at risk of being involved in gun violence. He assembled an interdisciplinary group of

JON LOWENSTEIN/NOOR/EYEVINE

Gang-related graffiti in Chicago, where around 3,500 people were shot last year — 246 were minors.

researchers who use artificial-intelligence (AI) techniques to study the language and images in social-media posts to identify patterns of grieving and anger.

By developing tools to automatically recognize these telltale emotional signs, Patton hopes to provide a way for community organizations to intervene before digital fights turn deadly. Programmes in Chicago are starting to take notice. "Our violence-prevention outreach has to change because gangs have changed," says Eddie Bocanegra, senior director of READI Chicago, an initiative aimed at reducing gun violence.

GUN VIOLENCE AND SOCIAL MEDIA

The United States leads the world in gun violence against children. About 1,300 minors die in shootings each year and another 5,800 are injured, according to researchers at the US Centers for Disease Control and Prevention[2]. Gun-related injuries are the third leading cause of death for children aged 17 and younger. And the death rate for African American children is ten times higher than for white and Asian American children.

Patton and others have found that social media has exacerbated gun violence among young people and changed how gangs recruit members, conduct business and initiate violence. It has become an acute problem in Chicago, Patton says, where street gangs have splintered into small, unruly crews with younger members who often settle minor disputes — such as insults on social media — with deadly force.

Violence and grief are common themes in Barnes's online and offline history. Between December 2011 and her death in 2014, she tweeted nearly 27,000 times. She adopted Facebook and Twitter names that paid homage to slain friends, and she vowed vengeance on their killers.

Barnes started hanging around with a local street crew in her early teens. In early 2011, a friend, Shondale 'Tooka' Gregory, was killed by gunfire while waiting for a bus. He was 15 years old. The crew started to refer to its territory as 'Tookaville' in memoriam, and Barnes began using 'Tookaville'kirah' as her Facebook name.

Later that year, when a member of a rival gang, 20-year-old Odee Perry, was killed just blocks away, Internet chatter suggested that Barnes was the shooter. She neither confirmed nor denied the speculation, which probably enhanced her online mystique.

In a survey of young black people in Chicago, nearly half reported that they had witnessed a gang-related killing[3]. Such experiences are especially detrimental to the adolescent brain, which is still developing, says Karen Sheehan, a paediatric emergency-room physician and professor at Lurie Children's Hospital and Northwestern University's Feinberg School of Medicine in Chicago. It affects the frontal lobe, she says, and diminishes the capacity to make good decisions. And the stress of grieving can exacerbate these deficiencies.

In 2012, Barnes experienced another tough loss. She witnessed first-hand as Tyquan Tyler, a 13-year-old boy and close friend, was killed by a stray bullet at a neighbourhood party. Barnes changed her Twitter username to 'TyquanAssassin' in memoriam, and her activity on the platform increased considerably.

Patton was most interested in the posts Barnes made in the days leading up to her own death. On 28 March 2014, 19-year-old Raason 'Lil B' Shaw, a member of her crew, was killed by Chicago police after he allegedly pointed a handgun at them during a foot chase. It was on the same street where Tyler had died, and Barnes was soon expressing her profound grief online. She renamed her twitter profile 'No Surrender Lil B' and tweeted darkly, presciently, "In da end we DIE," on 10 April. The very next day, Barnes was dead.

> "OUR RESEARCH HOPES TO SHOW THERE ARE PATHWAYS TO VIOLENCE."

GRIEF AND LOSS

The deaths of Shaw and Barnes helped motivate Patton to develop SAFElab, a research initiative that explores how urban youth of colour navigate their lives and express violence across social media. Patton and his team looked closely at 2,256 tweets posted in the weeks surrounding the killings from 235 unique users, all interacting with Barnes and her network[4]. Patton's goal was to sort tweets into categories. There were those expressing loss — including expressions of grief or sadness and descriptions of trauma, death or incarceration. And there were others signalling aggression — insults and threats of violence or the desire for revenge. Tweets that did not include features of aggression or loss were assigned to a catch-all category.

What the researchers found was a regular pattern of grief leading to aggression, often as rivals taunted those in pain or used language or imagery meant to minimize the loss of a friend or insult the dead. "They're going through a grieving process, and part of that process is anger and disbelief," says Patton. "Rival crews or gangs are interrupting the process. That heightens the conversation over time."

For example, one member of Barnes's crew tweeted shortly after Shaw's death, "Opps in trouble, we want blood." Shaw was killed by police officers, but his fellow crew members were enraged and wanted to take out their anger on an opposition, or 'opps', crew. A rival tweeted back, "come on our block mfs we want dis shit." (Original postings have been altered by SAFElab to protect the identity of the users.)

Patton's team found that expressions of grief significantly correlated with subsequent aggressive tweets, which happened about two days later, on average. Those two days could be a crucial window for a virtual or clinical intervention by counsellors, social workers or mental-health professionals, the researchers think. But combing through masses of Twitter traffic to detect and identify such signs of danger would be impossible. So Patton asked Columbia's Department of Computer Sciences to explore automating the search. The challenge caught the attention of Kathleen McKeown, a natural-language-processing specialist and founding director of the Data Science Institute at Columbia University in New York City.

Internet banging has its own distinctive lingua franca, which Patton describes as "a combination of African American vernacular English, social-media speak and brilliant uses of punctuation and numbers".

It was "very different than any language that I had ever worked on", says McKeown. To assist the computational interpretation, the team brought in violence specialists — individuals formerly involved with gangs — to annotate hundreds of tweets. Some of the translation was fairly obvious. A gun emoji and a devil face, for example, represent potential threats, and a phrase like "Got da Smithy on Me Right Na" would mean "I'm carrying a Smith & Wesson handgun." Other phrases were more opaque. "Keta smoking and thinkin of Lil B," was a tweet mourning the loss of Shaw, but also meant to disrespect a rival gang member who had been killed (see 'Tweet translations').

McKeown's group created a tool that could translate words and phrases. "We developed a bilingual dictionary," she says.

The first test runs of the tool accurately identified expressions of loss and grief 62% of the time, says Patton. By the time they published the research, this measure had improved. "We'll never be at 100% because social media is always changing and the language is often new. But this is very high for data science," Patton says.

They expanded the data set to include about 2 million tweets among 9,000 users, and the team is now training an additional AI application on the images that accompany posts. The researchers want to develop an AI application that can accurately identify grief and sadness even from facial expressions.

"The dominant narrative is that violence is a one-time incident like a finite moment in time. Our research hopes to show there are pathways to violence," says William R. Frey, a doctoral student who is mentored by Patton and is the coordinator of SAFElab.

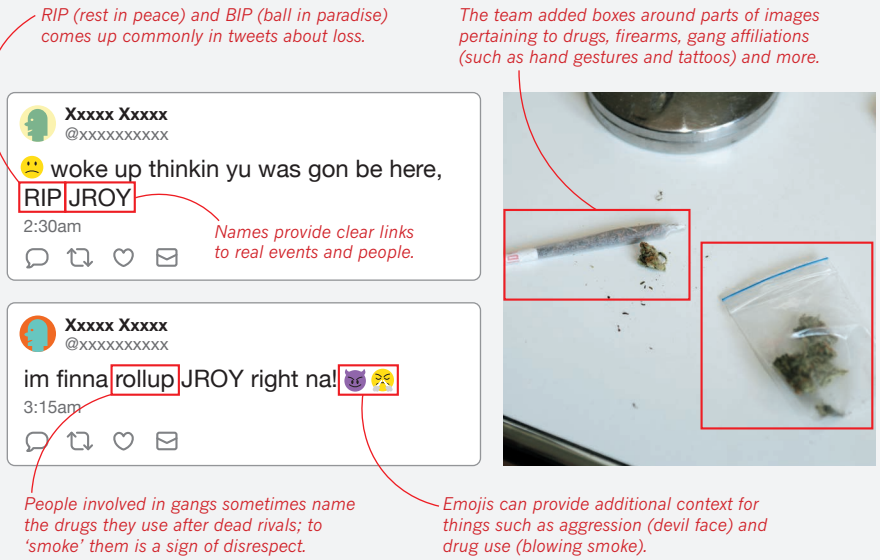
Research suggests that grief and grievances

TWEET TRANSLATIONS

SAFElab works with community organizations and people formerly involved with gangs to evaluate and contextualize tweets and images posted by young people in Chicago, Illinois.

ANNOTATION AND LABELLING

The team marked up thousands of tweets and images and then labelled them as pertaining to loss, aggression or other themes.



MACHINE LEARNING

Labelled tweets are used to train natural-language-processing and computer-vision models, which sort novel tweets into categories such as loss and aggression.

are data points along the road to violence, says Gene Deisinger, a psychologist, retired police officer and threat-management expert who has advised organizations and government agencies on how to identify and manage risks of violent behaviour. "We've learned that when a person fixates [on] the need for a violent resolution, that does increase the risk of violence."

Deisinger sees some similarities between Patton's research and threat-assessment activities. "You can actually develop some robust models of prediction of human behaviour such as what proportion of the group will escalate to violent behaviour. It still begs the question of what any individual will do," says Deisinger. He cautions against using group outcomes and predictions, particularly where taking action might interfere with constitutional rights.

EARLY INTERVENTIONS

Chicago has proved an ideal setting for Patton's experiments; the city has become a laboratory for strategies aiming to prevent injuries and deaths from gunshots. Cure Violence, a programme founded by epidemiologist Gary Slutkin at the University of Illinois at Chicago, aims to treat gun violence as if it were an infectious disease. It monitors activity and intervenes to interrupt the spread of gun violence, treats high-risk individuals and educates the community about prevention. READI Chicago offers employment opportunities, cognitive behavioural therapy and support services for young men at highest risk of violence.

One initiative that began in 2016 is the Institute for Nonviolence Chicago (INV), which was founded by Teny Gross, formerly of the Israel Defense Forces. He is aiming to partner with SAFElab to conduct field research on some of the intervention tools that Patton's group is creating.

INV has a hyper-local approach that organizes the community around the non-violent principles advocated by Martin Luther King Jr. Its 25 outreach workers follow chatter on social media, but they have a limited reach, says David Cassel, INV's director of strategy and organizational advancement. "They see conflicts happening in social media and then see the actual gun violence in the community. But it's very difficult for them to monitor social media and do their jobs," says Cassel. The organization hopes to get funding to test strategies for using SAFElab's tools, possibly to send automatic alerts to case workers, who could then contact people at risk of carrying out retaliatory attacks.

Patton's SAFElab project joins a growing body of research that uses AI and social data to predict public-health outcomes. An application called nEmesis, for example, developed by the University of Rochester in New York, also uses AI and natural-language processing to search Twitter. This program looks for tweets about food poisoning to help identify the source of an outbreak for public-health investigators.

The University of California Institute for Prediction Technology, based in Los Angeles, has been working on another problem. Scientists have used a statistical model and machine learning to predict heroin overdoses on the basis of Google searches for prescription and non-prescription opioids[5].

The study's lead author, Sean Young, says that Patton's research is "very promising", but he suggests there are some limitations. "How do we get a sense of the validity of the information?" he asks. How does one separate real threats from boastful swagger?

Patton acknowledges there is a degree of showmanship on social media. "People who live in communities impacted by violence may present themselves as being in a gang, but they are simply acting or performing," he says.

As SAFElab and INV look to secure funding for developing and testing tools for the real world, Patton and his colleagues hope that they can learn more from the short life of Gakirah Barnes and others like her. They have several new projects, he says, "including expanding Twitter analysis beyond aggression and grief to look at other factors associated with youth gun violence, such as substance use and mental health". Members of the team are now using network analysis to look at the links between online chatter and real-world behaviour. They are also experimenting with virtual reality to teach young people how to navigate social media and limit their exposure to violence.

Patton worries that some people might try to use social-media data in a discriminatory way by improperly trying to predict problems and even police activity in marginalized communities. But he sees many more positive aspects to the work: "I think the power of social media is that it gives you depth, vulnerability and multiple perspectives." He sees young people relating to each other and seeking out comfort and support. Strengthening positive connections in the community could be important for preventing the kind of violence that has ended so many young lives. He also hopes to broaden the discussion beyond violence. "With Gakirah, of course, we found a young woman who loved, who hurt, who was excited and who was in pain. That's a normal person, not just a gang member." ■

PHOTO: GETTY

Rod McCullom *is a science journalist in Chicago, Illinois.*

1. Patton, D. U., Eschmann, R. D. & Butler, D. A. *Comput. Human Behav.* **29**, A54–A59 (2013).
2. Fowler, K. A., Dahlberg, L. L., Haileyesus, T., Gutierrez, C. & Bacon, S. *Pediatrics* **140**, e20163486 (2017).
3. Voisin, D. R. & Neilands, T. B. *J. Adolesc. Health* **46**, 600–602 (2010).
4. Patton, D. U., Rambow, O., Auerbach, J., Li, K. & Frey, W. *NPJ Digit. Med.* **1**, 11 (2018).
5. Young, S. D., Zheng, K., Chu, L. F. & Humphreys, K. *Drug Alcohol Depend.* **190**, 166–169 (2018).

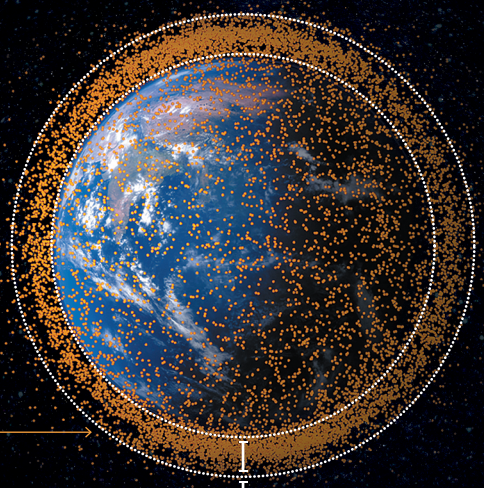
THE QUEST TO CONQUER THE SPACE JUNK PROBLEM

Zombie satellites, rocket shards and collision debris are creating major traffic risks in orbits around Earth. Researchers are working to reduce the threats posed by more than 20,000 objects in space.

BY ALEXANDRA WITZE

BUSY SKIES

There are currently more than 20,000 objects in orbit around Earth, according to catalogues that track operational satellites, dead ones and other human-made debris, such as pieces from rockets. And the problem is growing quickly: more than 1,800 new objects joined the crowded skies in 2017.

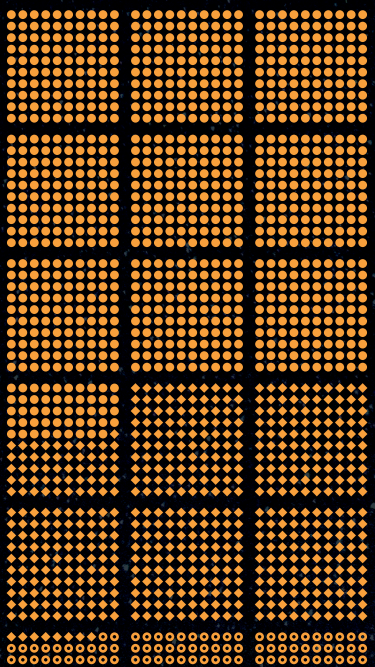


A visualization by NASA depicts the traffic of objects in orbits around Earth.

TYPE OF DEBRIS

- ● = 10 × Payload related
- ◆ = 10 × Rocket related
- ○ = 10 × Unknown

LOW EARTH ORBIT (LEO):
altitudes up to 2,000 kilometres
Not all objects in this count are confined to low Earth orbit. Some pass through LEO and travel farther from the planet.



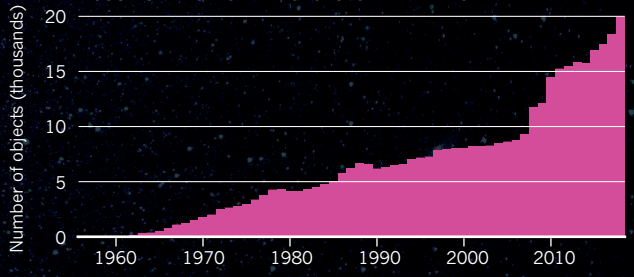
15,900 objects

MEDIUM EARTH ORBIT (MEO)
altitudes between 2,000 and 35,000 km

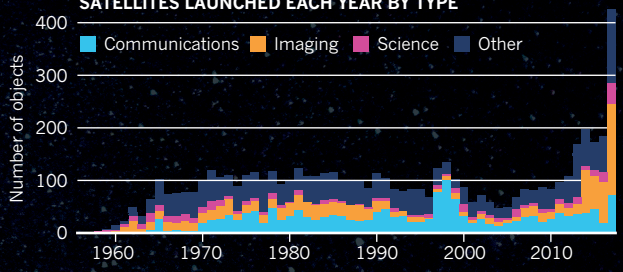
GEOSTATIONARY ORBIT (GEO)
altitudes around 35,000 km
Used for some communications and weather satellites. This count includes objects that pass through GEO orbits.

2,931 objects

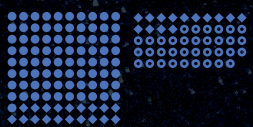
ACCUMULATION OF ALL OBJECTS IN EARTH ORBIT



SATELLITES LAUNCHED EACH YEAR BY TYPE



OTHER ORBITS
Includes medium Earth orbits



1,488 objects

On Monday 2 July, the CryoSat-2 spacecraft was orbiting as usual, just over 700 kilometres above Earth's surface. But that day, mission controllers at the European Space Agency (ESA) realized they had a problem: a piece of space debris was hurtling uncontrollably towards the €140-million (US$162-million) satellite, which monitors ice on the planet.

As engineers tracked the paths of both objects, the chances of a collision slowly increased — forcing mission controllers to take action. On 9 July, ESA fired the thrusters on CryoSat-2 to boost it into a higher orbit. Just 50 minutes later, the debris rocketed past at 4.1 kilometres a second.

This kind of manoeuvre is becoming much more common each year, as space around Earth grows increasingly congested. In 2017, commercial companies, military and civil departments and amateurs lofted more than 400 satellites into orbit, over 4 times the yearly average for 2000–2010. Numbers could rise even more sharply if companies such as Boeing, OneWeb and SpaceX follow through on plans to deploy hundreds to thousands of communications satellites into space in the next few years. If all these proposed 'megaconstellations' go up, they will roughly equal the number of satellites that humanity has launched in the history of spaceflight.

All that traffic can lead to disaster. In 2009, a US commercial Iridium satellite smashed into an inactive Russian communications satellite called Cosmos-2251, creating thousands of new pieces of space shrapnel that now threaten other satellites in low Earth orbit — the zone stretching up to 2,000 kilometres in altitude. Altogether, there are roughly 20,000 human-made objects in orbit, from working satellites to small shards of solar panels and rocket pieces. And satellite operators can't steer away from all potential collisions, because each move consumes time and fuel that could otherwise be used for the spacecraft's main job.

"IF WE GO ON LIKE THIS, WE WILL REACH A POINT OF NO RETURN."

Concern about space junk goes back to the beginning of the satellite era, but the number of objects in orbit is rising so rapidly that researchers are investigating new ways of attacking the problem. Several teams are trying to improve methods for assessing what is in orbit, so that satellite operators can work more efficiently in ever-more-crowded space. Some researchers are now starting to compile a massive data set that includes the best possible information on where everything is in orbit. Others are developing taxonomies of space junk — working out how to measure properties such as the shape and size of an object, so that satellite operators know how much to worry about what's coming their way. And several investigators are identifying special orbits that satellites could be moved into after they finish their missions so they burn up in the atmosphere quickly, helping to clean up space.

The alternative, many say, is unthinkable. Just a few uncontrolled space crashes could generate enough debris to set off a runaway cascade of fragments, rendering near-Earth space unusable. "If we go on like this, we will reach a point of no return," says Carolin Frueh, an astrodynamical researcher at Purdue University in West Lafayette, Indiana.

DIRTYING ORBITS

Astronomers and others have worried about space junk since the 1960s, when they argued against a US military project that would send millions of small copper needles into orbit. The needles were meant to enable radio communications if high-altitude nuclear testing were to wipe out the ionosphere, the atmospheric layer that reflects radio waves over long distances. The Air Force sent the needles into orbit in 1963, where they successfully formed a reflective belt. Most of the needles fell naturally out of orbit over the next three years, but concern over 'dirtying' space nevertheless helped to end the project.

It was one of the first examples of the public viewing space as a landscape that should be kept clean, says Lisa Rand, a historian of science in Philadelphia, Pennsylvania, and a fellow with the American Historical Association and NASA.

Since the Soviet Union launched the first satellite, Sputnik, in 1957, the number of objects in space has surged, reaching roughly 2,000 in 1970, about 7,500 in 2000 and about 20,000 known items today. The two biggest spikes in orbital debris came in 2007, when the Chinese government blew up one of its satellites in a missile test, and in the 2009 Iridium–Cosmos collision. Both events generated thousands of fresh fragments, and they account for about half of the 20-plus satellite maneouvres that ESA conducts each year, says Holger Krag, head of ESA's space-debris office in Darmstadt, Germany.

Each day, the US military issues an average of 21 warnings of potential space collisions. Those numbers are likely to rise dramatically next year, when the Air Force switches on a powerful new radar facility on Kwajalein in the Pacific Ocean. That facility will allow the US military to detect objects smaller than today's 10-centimetre limit for low Earth orbit, and this could increase the number of tracked objects by a factor of five.

Even as our ability to monitor space objects increases, so too does the total number of items in orbit. That means companies, governments and other players in space are having to collaborate in new ways to avoid a shared threat. Since the 2000s, international groups such as the Inter-Agency Space Debris Coordination Committee have developed guidelines for achieving space sustainability. Those include inactivating satellites at the end of their useful lifetimes by venting leftover fuel or other pressurized materials that could lead to explosions. The intergovernmental groups also recommend lowering satellites deep enough into the atmosphere that they will burn up or disintegrate within 25 years.

But so far, only about half of all missions have abided by this 25-year guideline, says Krag. Operators of the planned megaconstellations say they will be responsible stewards of space, but Krag worries that the problem could increase, despite their best intentions. "What happens to those that fail or go bankrupt?" he asks. "They are probably not going to spend money to remove their satellites from space."

TRAFFIC COPS FOR SPACE

In theory, satellite operators should have plenty of room for all these missions to fly safely without ever nearing another object. So some scientists are tackling the problem of space junk by trying to understand where all the debris is to a high degree of precision. That would alleviate the need for many unnecessary manoeuvres that today are used to avoid potential collisions. "If you knew exactly where everything was, you would almost never have a problem," says Marlon Sorge, a space-debris specialist at the Aerospace Corporation in El Segundo, California.

The field is called space-traffic management, because it's analogous to managing traffic on the roads or in the air. Think about a busy day at an airport, says Moriba Jah, an astrodynamicist at the University of Texas at Austin: planes line up in the sky like a string of pearls, landing and taking off close to one another in a carefully choreographed routine. Air-traffic controllers know the location of the planes down to 1 metre in accuracy.

The same can't be said for space debris. Not all objects in orbit are known, and even those included in databases are tracked to varying levels of precision. On top of that, there is no authoritative catalogue that accurately lists the orbits of all known space debris.

Jah illustrates this with a web-based database that he developed, called ASTRIAGraph. It draws on several sources, such as catalogues maintained by the US and Russian goverments, to visualize the locations of objects in space. When he types in an identifier for a particular space object, ASTRIAGraph draws a purple line to designate its orbit.

Only this doesn't quite work for a number of objects, such as a Russian rocket body launched in 2007 and designated in the database as object number 32280. When Jah enters that number, ASTRIAGraph draws two purple lines: the US and Russian sources contain two completely different orbits for the same object. Jah says that it is almost impossible to tell which is correct, unless a third

SOURCE: ESA ANNUAL SPACE ENVIRONMENT REPORT. EARTH DEBRIS IMAGE: NASA GODDARD SPACE FLIGHT CENTER/JSC

source of information could help to cross-correlate the correct location.

ASTRIAGraph currently contains some, but not all, of the major sources of information about tracking space objects. The US military catalogue — the largest such database publicly available — almost certainly omits information on classified satellites. The Russian government similarly holds many of its data close. Several commercial space-tracking databases have sprung up in the past few years, and most of those do not share openly.

Jah describes himself as a space environmentalist: "I want to make space a place that is safe to operate, that is free and useful for future generations." Until that happens, he argues, the space community will continue devolving into a tragedy of the commons, in which all space-flight operators are polluting a common resource.

He and other space environmentalists are starting to make headway, at least when it comes to US space policy. Jah testified on space-traffic management in front of Congress last year, at the invitation of Ted Cruz, a Republican senator from Texas who co-introduced a space-regulations bill this July. In June, President Donald Trump also signed a directive on space policy that, among other things, would shift responsibility for the US public space-debris catalogue from the military to a civilian agency — probably the Department of Commerce, which regulates business.

NASA

Tiny CubeSats are released from the International Space Station in 2012.

The space-policy directive is a rare opportunity to discuss space junk at the highest levels of the US government. "This is the first time we're really having this conversation in a serious fashion," says Mike Gold, vice-president for regulatory, policy and government contracts at Maxar Technologies of Westminster, Colorado, which owns and operates a number of satellites.

THE ORBITING DEAD

The space around Earth is filled with zombies: some 95% of all objects in orbit are dead satellites or pieces of inactive ones. When someone operating an active satellite gets an alert about an object on a collision course, it would be helpful to know how dangerous that incoming debris is. "With more and more objects, and the uncertainties we currently have, you just get collision warnings no end," says Frueh. (Micrometeorites represent a separate threat and can't be tracked at all.)

To assess the risk of an impending collision, satellite operators need to know what the object is, but tracking catalogues have little information about many items. In those cases, the military and other space trackers use telescopes to gather clues in the short period before a potential collision.

Working with the Air Force, Frueh and her colleagues are developing methods to rapidly decipher details of orbiting objects even when very little is known about them. By studying how an object reflects sunlight as it passes overhead, for instance, she can deduce whether it is tumbling or stable — a clue to whether or not it is operational. Her team is also experimenting with a machine-learning algorithm that could speed up the process of characterizing items, work she will describe on 14 September at a space-tracking meeting in Maui, Hawaii.

Once researchers know what an orbiting object is made of, they have a number of potential ways to reduce its threat. Some sci-fi-tinged proposals involve using magnets to sweep up space junk, or lasers to obliterate or deflect debris in orbit. In the coming weeks, researchers at the University of Surrey in Guildford, UK, will experiment with a net to ensnare a test satellite. The project, called RemoveDEBRIS, will then redirect the satellite into an orbit that will re-enter the atmosphere.

But such active approaches to cleaning up space junk aren't likely to be practical over the long term, given the huge number of objects in orbit. So some other experts consider the best way of mitigating space junk to be a passive approach. This takes advantage of the gravitational pulls of the Sun and the Moon, known as resonances, that can put the satellites on a path to destruction. At the University of Arizona in Tucson, astrodynamicist Aaron Rosengren is developing ways to do so.

Rosengren first came across the idea when studying the fates of satellites in medium Earth orbit (MEO). These travel at altitudes anywhere between about 2,000 kilometres up, where low Earth orbit ends, and 35,000 kilometres up, where geostationary orbits begin.

Satellites in low Earth orbit can be disposed of by forcing them to re-enter the atmosphere, and most satellites in the less heavily trafficked geostationary region can be safely placed in 'graveyard' orbits that never interact with other objects. But in MEO, satellite trajectories can be unstable over the long term because of gravitational resonances.

An early hint that spacecraft operators could harness this phenomenon came from ESA's INTEGRAL γ-ray space telescope, which launched in 2002. INTEGRAL travels in a stretched-out orbit that spans all the way from low Earth orbit, through MEO, and into geostationary orbit. It would normally have remained in space for more than a century, but in 2015, ESA decided to tweak its orbit. With a few small thruster burns, mission controllers placed it on a path to interact with gravitational resonances. It will now re-enter the atmosphere in 2029, rather than decades later.

In 2016, Rosengren and his colleagues in France and Italy showed that there is a dense web of orbital resonances that dictates how objects behave in MEO (J. Daquin *et al. Celest. Mech. Dyn. Astr.* **124**, 335–366; 2016). Rosengren thinks this might offer a potential solution. There are paths in this web of resonances that lead not to MEO, but directly into the atmosphere, and operators could take advantage of them to send satellites straight to their doom. "We call it passive disposal through resonances and instabilities," says Rosengren. "Yeah, we need a new name."

Other researchers have explored the concept before, but Rosengren is trying to push it into the mainstream. "It's one of the newer things in space debris," he says.

These disposal highways in the sky could be easy to access. At a space conference in July in Pasadena, California, Rosengren and his colleagues reported on their analysis of US Orbiting Geophysical Observatory satellites from the 1960s. The scientists found that changing the launch date or time by as little as 15 minutes could lead to huge differences in how long a satellite remains in orbit. Such information could be used to help calculate the best times to depart the launch pad.

Being proactive now could head off a lot of trouble down the road, as operators of satellites such as CryoSat-2 have found. When ESA decided to take evasive action in early July, its engineers had to scramble and work through the weekend to get ready for the manoeuvre. Once the space junk had safely flown by, CryoSat-2 took a few days to get back into its normal orbit, says Vitali Braun, a space-debris engineer with ESA.

But the alerts didn't stop coming. In the weeks that followed, mission controllers had to shift various satellites at least six times to dodge debris. And on 23 August, they nudged the Sentinel-3B satellite out of the way of space junk for the first time. It had been in orbit for only four months. ■

Alexandra Witze *is a correspondent for* Nature *based in Boulder, Colorado.*

COMMENT

UESLEI MARCELINO/REUTERS

Smog envelopes Santiago in Chile.

Five steps to improve air-quality forecasts

A worldwide monitoring and modelling network would reduce the dramatic toll of air pollution on health and food production, urge **Rajesh Kumar** and colleagues.

Seven million people die every year from the effects of air pollution. More than 90% of such deaths are in developing countries[1]. Across southern Asia, levels of fine particulate matter ($PM_{2.5}$) and surface ozone exceed the World Health Organization (WHO) limits for much of the year[2]. Ozone damage to crops and plants — especially to soya beans, wheat and maize (corn) — results in 79 million to 121 million tonnes of lost produce globally, at a cost of US$11 billion to $18 billion[3]. India's crop losses alone would feed 94 million people[4]. All this costs the world's economy US$5 trillion per year[5].

But air pollution often goes unmonitored. Some of the fastest-growing cities in Africa, including Lagos, Kinshasa, Abidjan and Dakar, have no air-quality alert systems. Governments can be reluctant to acknowledge the problem, or lack the tools to address it. There is no international strategy for dealing with the issue. And few people are trained in how to collect and interpret air-quality data.

Improvements can take decades. It took the US city of Atlanta, Georgia, 15 years to reduce emissions from power plants by around 80% and from traffic by up to 90%, avoiding more than 50,000 hospital visits for asthma and lung diseases (ref. 6). Los Angeles in California took 50 years to reduce ozone levels by two-thirds[7].

Forecasts of hazardous air pollution are crucial to help reduce exposure. Vulnerable people can avoid strenuous outdoor activities or stay indoors. Schools might restrict outdoor sports activities, parents can limit the time their children spend outdoors and doctors might advise their patients to stay inside when levels are high. In Canada, ▶

for example, daily forecasts have helped to decrease the number of asthma-related emergency-department visits by 25% (ref. 8).

City managers can also take steps to limit pollution. For example, Santiago in Chile restricts driving and certain industries during predicted periods of high pollution. These measures have reduced the $PM_{2.5}$ concentrations by 20% and avoided around 8 deaths each day[9]. However, efforts to ban cars on polluted days in Delhi have had little effect[10].

Many parts of North America and Europe provide daily broadcasts of air-quality forecasts. But across much of Asia, Africa and South America, smog still arrives unannounced. Predictions require advanced computer models and regional weather forecasts; both are lacking across the developing world.

Access to such systems needs to be made global. Local air quality is affected by the intercontinental transport of air pollutants and distant weather events such as the El Niño Southern Oscillation (ENSO). Denser sampling of pollution data would tell air-quality managers which activities are most dangerous and where the worst pollution is coming from — road traffic or industry, say.

Most forecasts of air quality cover two to five days. Extending this to seven to ten days — similar to alert periods for hurricanes and floods — would give communities and hospitals more time to prepare. Seasonal forecasts would give farmers the chance to shift their planting and harvesting dates and choose resistant crops. Such extensions, as well as predictions that span seasons, will be helped by the ongoing efforts to improve weather forecasts and predict emissions from wildfires.

Rolling out a global forecasting system for air quality will take five advances: expanding observing networks; improving models; devising metrics and tools for quantifying air pollution; disseminating the information; and training experts. These steps align with the priorities for weather and climate set out by the World Meteorological Organization (WMO)[11].

We call on the WMO, the WHO, the United Nations Environment Programme (UNEP) and the Food and Agriculture Organization of the UN (FAO) to lead the development of an international programme for air-quality monitoring and prediction. Financial institutions such as the World Bank and non-governmental organizations should support air-quality initiatives.

FIVE STEPS

Monitoring. Levels of ozone, $PM_{2.5}$, carbon monoxide, nitrogen oxides, sulfur dioxide, aerosols and other pollutants need to be tracked globally, at least daily and ideally hourly. Satellite, aircraft, balloon and ground-based instruments will all be needed. Upcoming geostationary missions such as the US Tropospheric Emissions: Monitoring of Pollutants; the Korean Aerospace Research Institute (KARI) Geostationary Environmental Monitoring Spectrometer; and the European Copernicus Programme's Sentinel-4 instruments will soon be able to track pollutants continuously over North America, Asia and Europe. Governments should launch more such satellites or drones to cover other continents. The costs (of $50 million to $100 million) are small relative to the economic losses from air pollution.

On the ground, air-quality monitoring networks need to be established in Africa, South America and southeast Asia. With about 1 station per urban neighbourhood and around 100 stations providing a baseline outside cities, these could complement the hundreds of Global Earth Observatory stations that have been proposed for tracking many environmental compounds[12]. Local scientists will need to be trained to gather the data, with support from governments and multilateral exchange programmes for researchers.

"Rolling out a global forecasting system for air quality will take five advances."

All countries should share their air-quality information. Some do already; many do not. There is no international authority that organizes this. The WMO facilitates such sharing for weather data, and it should coordinate an international agreement and devise quality controls and operating guidelines for air quality, too. Metadata should be included (such as the height of a detector above ground, or its distance from the nearest building, tree, road or industrial stack) as well as measurement errors.

Modelling. Researchers must improve regional and urban air-quality prediction models. These combine weather and atmospheric composition, emissions records and ground and space observations. The models should pinpoint major sources of pollution in cities, such as buildings or roads. And they should track daily cycles, for example in pollution from traffic, as well as decadal changes in weather patterns and emissions. Natural emissions, including those from wildfires, must be accounted for.

Uncertainties must be quantified to avoid false alarms or missed episodes. Inventories that collate emissions from various sources (such as residential and industrial buildings, power, traffic, agriculture and forest fires) are a major source of error, especially in developing countries where reporting is sparse. Such data are often years out of date. Biased measurements and poorly understood physical and chemical processes (such as organic chemistry in the atmosphere, aerosol microphysics and deposition mechanisms) also introduce uncertainties.

Many developing countries do not have the computing power to model air quality. This capability should be added to existing weather-forecasting systems. Large institutions with advanced supercomputers should run the global forecasts, such as the European Centre for Medium-Range Weather Forecasts (ECMWF) and the US National Center for Atmospheric Research (NCAR). National meteorological, air-quality and environmental services must tailor models to local needs. Cloud computing can broaden access, and the ECMWF is trialling this for its weather forecasts.

Interpretation. Researchers need to understand the processes that control air pollution in particular places and over the longer term. They should develop tools to help policymakers and city managers evaluate

SAJJAD HUSSAIN/AFP/GETTY

Schoolchildren walk through smog in New Delhi.

MIKKEL OSTERGAARD/PANOS

An air-pollution monitoring device in Nepal.

the impacts of regulations, economic policies and urbanization on air quality.

Broad collaborations will be needed. For example, the Partnership with China on Space Data (PANDA) project is bringing together air-quality scientists and officials from Europe and China to develop warning systems for 28 Chinese cities. Similar partnerships are springing up in South America and India, but still need to be developed for Africa and southeast Asia.

Dissemination. National meteorological, air-quality and environmental service providers should calculate health and air-quality indices based on pollutant concentrations, similar to those used in North America and Europe. For instance, the Air Quality Index published by the US Environmental Protection Agency reports how polluted the air is (on a scale of 0–500) and the health effects (such as unhealthy for sensitive groups, or hazardous for everyone). These indices should be explained and released to the public and to hospitals through websites, mobile messaging and so on. London, UK, and Riga, Latvia, issue text alerts on air quality, for example. Air quality could be added to other alert apps, such as one indicating risks from lightning strikes in Andhra Pradesh, India. Technology companies are showing more interest in air-quality monitoring and forecasting. Google, for instance, has been mapping air quality since 2014 with its street-view cars, and IBM has launched a global initiative called Green Horizons, in which air-quality forecasting is a key component.

Training and education. Students, environmental engineers and early-career scientists in developing countries need training in how to measure air quality and climate change and interpret the results. Exchanges of scientists between developed and developing countries could be arranged through programmes such as the European Commission's Marie Skłodowska-Curie Actions. Massive open online courses are a great opportunity to gain detailed first hand insight. In October, for instance, the European Organisation for the Exploitation of Meteorological Satellites and the ECMWF are running a course on the free platform Iversity, called Monitoring Air Quality from Space, which will explain more about the data being gathered and delivered in the Copernicus Programme. The public and authorities need to understand how to change behaviours, for example by avoiding driving when pollution is bad. Incentives will be needed, such as subsidies for people using public transport.

As a first step, a summit sponsored by the WMO, WHO, UNEP and FAO could begin to design a global strategy for reducing deaths caused by air pollution. ■

Rajesh Kumar *is a project scientist at the National Center for Atmospheric Research (NCAR), Boulder, Colorado, USA.* **Vincent-Henri Peuch** *heads the Copernicus Atmosphere Monitoring Service operated by the European Center for Medium-Range Weather Forecasts in Reading, UK.* **James H. Crawford** *is NASA's senior scientist for tropospheric chemistry at the Langley Research Center in Hampton, Virginia, USA.* **Guy Brasseur** *is director of the Atmospheric Chemistry Observations and Modeling Laboratory at NCAR.*
e-mail: rkumar@ucar.edu

1. World Health Organization. *Air pollution*; available at https://go.nature.com/2bzdas7
2. Kumar, R. *et al. J. Geophys. Res. Atmos.* **123**, 1840–1864 (2018).
3. Avnery, S., Mauzerall, D. L., Liu, J. & Horowitz, L. W. *Atmos. Environ.* **45**, 2284–2296, (2011).
4. Ghude, S. D. *et al. Geophys. Res. Lett.* **41**, 5685–5691 (2014).
5. World Bank & Institute for Health Metrics and Evaluation. *The Cost of Air Pollution: Strengthening the Economic Case for Action* (International Bank for Reconstruction and Development/World Bank, 2016); available at https://go.nature.com/2mluupw
6. Russell, A. G. *et al. Impacts of Regulations on Air Quality and Emergency Department Visits in the Atlanta Metropolitan Area, 1999–2013* HEI Research Report 195 (Health Effects Institute, 2018).
7. Parrish, D. D., Xu, J., Croes, B. & Shao, M. *Front. Environ. Sci. Eng.* **10,** 11 (2016).
8. Chen, H. *et al. Lancet Planet. Health* **2**, e19–e26 (2018).
9. Mullins, J. & Bharadwaj, P. *Am. J. Agric. Econ.* **97**, 1107–1134 (2015).
10. Chandra, B. P. *et al. Curr. Sci.* **114**, 1318–1325 (2018).
11. Hov, Ø. *et al. Nature* **552**, 168–170 (2017).
12. Kulmala, M. *Nature* **553**, 21–23 (2018).

A waterfall in Grand Canyon National Park, Arizona.

JACK DYKINGA/NATUREPL.COM

MATERIALS SCIENCE

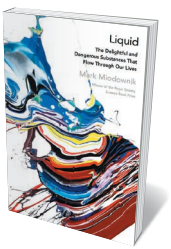
Wine, water, oil and spit

Derek Lowe enjoys Mark Miodownik's sparkling journey through liquids.

We humans can't help but have our views of reality skewed by our own experience. We take the physical conditions around us as the normal state of affairs and regard others as extreme. The chemist's 'standard temperature and pressure' of 0°C under 10^5 pascals makes perfect sense as a reference in the human frame, but is quite unusual in the larger scheme of things. The rest of the Solar System, for example, often features temperatures and pressures much higher or much lower; the rest of (mostly empty) outer space is even worse. It's the surface of Earth that is the outlier.

These uncommon conditions mean we experience something very rare in the higher and lower expanses of temperature and pressure: a wide variety of liquids. It is this odd realm that materials scientist Mark Miodownik explores in *Liquid*, his enjoyable successor to his 2013 paean to materials, *Stuff Matters*.

The book explores the histories, structures and properties of many different sorts of liquid, with excursions into the larger topics each brings up. Miodownik organizes his narrative around the conceit of an aeroplane journey, with various incidents and episodes setting off trains of thought. Alcohol makes an early appearance by way of the drinks trolley. Others are introduced through ocean waves observed below, the soap dispenser in the lavatory, the thought of refrigerants in the air-conditioning system, and so on. The framework is reasonably effective, although it gets a bit wearing, and a few digressions are surprisingly lengthy. (Then again, it is a transatlantic flight.)



Liquid: The Delightful and Dangerous Substances That Flow Through Our Lives
MARK MIODOWNIK
Viking (2018)

The liquid state is a small part of most phase diagrams, a narrow range between a large solid zone and a large gaseous one. Conditions have to be just right for a substance to condense out of the gas phase but not to firm up into some kind of solidified mass. And for many simple compounds, those liquid conditions manage to overlap with how we're used to seeing and handling them — which is why a plane journey can include so many useful examples. The one we all know best, of course, is water. The discovery of liquid water on another planet or moon, even far below the surface, somehow makes that world seem more real to us. But, as Miodownik explains, it's one of the oddest liquids of all.

H_2O is strangely sticky and viscous for a molecule with such a small molecular weight. It has abnormally high melting

and boiling points compared with anything chemically similar, such as ammonia or hydrogen sulfide. When it freezes, its solid phase is actually less dense than its liquid one. That relatively rare characteristic leads to ice cubes and icebergs floating instead of sinking as any normal solid phase should do. (If there are sentient creatures somewhere in the Universe living next to lakes of superfluid liquid helium, they would probably be confused by water, which flows with weird side effects of friction and drag unknown in the superfluid world.)

"The rest of the Solar System often features extremely high or low temperatures and pressures. It's the surface of Earth that is the outlier."

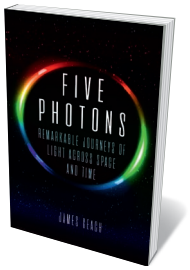
Liquid distils a great deal of interesting information in accurate, readable form. As a chemist, I find it a relief to read such an overview without being distracted by mischaracterized or oversimplified details. Solid-state physicists and materials scientists will also celebrate Miodownik's excellent efforts at tying the everyday properties of liquids to their molecular structure. He provides many vivid examples: among these are saliva (during the flight's meal service, naturally) and jet fuel, which he notes is not an explosive, but still has more chemical energy per unit volume than nitroglycerine.

One of the things that chemistry and physics teach is that the information given by our senses is only a small part of the story: water is wet, our fingers can tell us that much. What we don't feel are the water molecules themselves, interacting with the protein surfaces of our skin. Their very atoms and electron clouds come within range of each other, attracting and repelling and adding up to the sensations that our vivid (but often crude) senses interpret. That hidden world underlies every object we see and handle. *Liquid* gives readers a sense of this — no small feat.

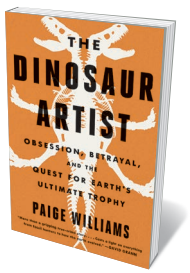
And that brings up the question of who might read it. As with *Stuff Matters*, Miodownik is inspiring those in search of science in an accessible, entertaining format. Today, materials scientists are preparing exotic fluids packed with nanoparticles that can turn them into magnets or optical sensors, and nanotechnologists and molecular biologists are exploring the behaviour of water and other liquids on very small scales. *Liquid* will come in very useful for people eager to understand these advances. ■

Derek Lowe *has worked in early-stage drug discovery for decades. His* In the Pipeline *is one of the longest-running science blogs.*
e-mail: derekb.lowe@gmail.com

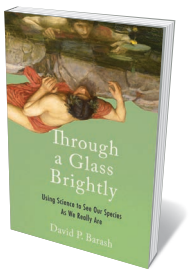
Books in brief



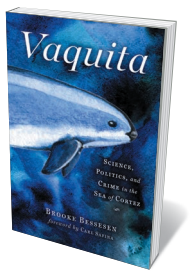
Five Photons
James Geach REAKTION *(2018)*
Light illuminates cosmic origins and decodes quotidian realities. But what is it? This deft primer by astrophysicist James Geach captures the elusive electromagnetic wave in five processes. His meditation on 'old' light takes us back to the singularity: the "cosmic seed" that expanded into the Big Bang. A study of starlight plunges us into the seething stellar surface. We peruse dark energy, radio waves and quasars — beacon-like galaxies in which supermassive black holes feed off interstellar gas and release vast amounts of energy. A masterclass in elucidating hard science with elegance and brevity.



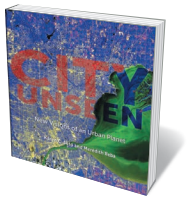
The Dinosaur Artist
Paige Williams HACHETTE *(2018)*
Who owns fossils? That vexed question lies at the heart of this exposé of the global trade in dinosaur remains — a messy meeting-place of commercial fossil collectors, palaeontologists, wealthy enthusiasts and natural-history museums. New Yorker staff writer Paige Williams's packed account centres on former Mongolian president Tsakhiagiin Elbegdorj, US dinosaur hunter and restorer Eric Prokopi and a costly *Tarbosaurus bataar* fossil. An astonishing tangle of financial gain, national identity, scientific fervour and, above all, the obsessional need to possess pieces of the past.



Through a Glass Brightly
David P. Barash OXFORD UNIVERSITY PRESS *(2018)*
As a species, we seem to be unable to shake off the idea of our exceptionalism. Yet science regularly trounces such ideas, argues evolutionary biologist David Barash in this briskly erudite study. Barash punctures human paradigms such as the 'anthropic principle', rationality and even selfhood, marshalling considerable research and considered reasoning as he goes. He concludes, rather splendidly, that the loss of such illusions flings open the door "to do something really extraordinary: to see ourselves as we really are" and use that knowledge to behave with more humanity.



Vaquita: Science, Politics, and Crime in the Sea of Cortez
Brooke Bessesen ISLAND *(2018)*
The world's smallest cetacean, the vaquita (*Phocoena sinus*), is also the most endangered marine mammal on the planet, found solely in northern Mexico's Gulf of California. In this intrepid conservation detective story, marine biologist Brooke Bessesen deconstructs the species' demise, showing how the tiny porpoises drown in gillnets used for poaching a prized black-market fish, *Totoaba macdonaldi*. As she shows, the effort to conserve remaining vaquitas is a torturously uncertain challenge — but ever driven by the idea, articulated by field biologist George Schaller, that "we cannot recover a lost world".



City Unseen: New Visions of an Urban Planet
Karen Seto and Meredith Reba YALE UNIVERSITY PRESS *(2018)*
Cities are a tug-of-war between nature and humanity — their configuration shaped by topography even as they mould the environment in and around them. This stunning study by Karen Seto and Meredith Reba explores this uneasy symbiosis through surreally hued satellite images of 100 cities. Snaps of Phoenix, Arizona, taken 31 years apart reveal serious urban sprawl, and a shot of grain fields around Semikarakorsk, Russia, is a controlled riot of colour and line with the verve of early modernist art. **Barbara Kiser**

STRUCTURAL BIOLOGY

From ribosome to Royal Society

Georgina Ferry enjoys Venki Ramakrishnan's account of his road to the Nobel Prize.

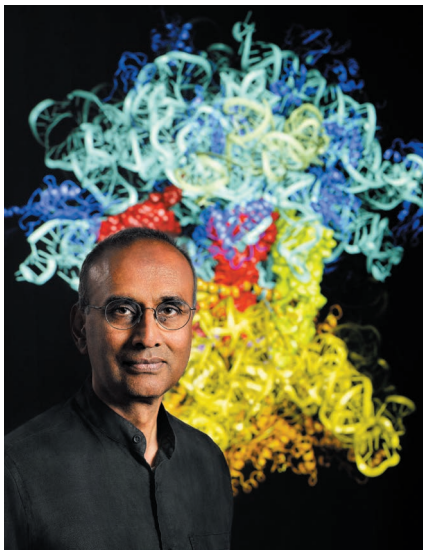
In 1968, the molecular biologist James Watson electrified the staid world of scientific biography by publishing *The Double Helix*, his memoir of the discovery of DNA's structure. Its irreverent style shocked and delighted readers in equal measure. And it changed how many think about science, by presenting it as a race in which winners got Nobel prizes and losers got nothing.

Fifty years later, structural biologist, Nobel laureate and Royal Society president Venki Ramakrishnan tells the story of his own marathon. In *Gene Machine*, he thoughtfully embeds his trajectory in a wider meditation on how scientists make the decisions that lead to success or failure — and on how they struggle to solve complex problems. "Scientists will collaborate or compete depending on what is in their self-interest," he writes, adding that competition can be "good for science, even if it isn't so great for scientists."

Ramakrishnan recognizes from the outset that, although most people today have at least a vague idea that DNA is the carrier of genetic information, the ribosome barely registers. Yet, without it, life as we know it would not have evolved. Every living cell contains millions of these complex structures, each consisting of a large and small subunit and incorporating a few dozen proteins and long strands of RNA. They are themselves protein factories. Each ribosome works along a molecule of messenger RNA, reading its copy of the genetic instructions encoded in DNA. Guided by this, ribosomes assemble protein chains that then float off to do the work of building and running the body.

This much was already known when Ramakrishnan began to work on the ribosome in 1978 — although, as he writes, "we had no idea how it did even one of the many complicated steps involved in making a protein". The race for the ribosome would reframe it as an intricate machine with numerous working parts — which you can now watch in action on YouTube.

Indian-born, with academic parents, Ramakrishnan studied physics at the University of Baroda (now Vadodara) in Gujarat. In 1971, he went to Ohio University in Athens for graduate work in theoretical physics. By the time he had submitted his thesis, he had decided to switch to biology. Reading around a subject about which he knew almost nothing, he picked the ribosome on the strength of a *Scientific American* article co-authored by biologist Peter Moore. Within a few years he was on the starting blocks, as a postdoc in Moore's laboratory at Yale University in New Haven, Connecticut.



MAX ALEXANDER/SPL

Venki Ramakrishnan.

It was clear that the only way to understand the ribosome was to solve its 3D structure at the resolution of individual atoms. Ramakrishnan retrained in X-ray crystallography, then the preferred technique of structural biologists.

Structural biologists face near-impossible challenges. These include making floppy, irregular proteins form crystals to bombard with X-rays; obtaining a diffraction pattern from the irradiated crystals before the structures disintegrate; and solving the ever-present 'phase problem', the ambiguity in a diffraction pattern caused by peaks in an X-ray wave giving the same intensity as troughs. Eventually, software generates maps of electron density at high-enough resolution to 'see' the individual atoms. For the big, complicated ribosome, many research teams in different countries worked on each of these problems successively or simultaneously.

The acknowledged founder of the ribosome field is Israeli crystallographer Ada Yonath. She, with colleagues at the Max Planck Institute for Molecular Genetics in Berlin, published the first successful crystallization of a ribosomal subunit in 1980. An unwritten rule in crystallography at the time was that once someone had a crystal, everyone else would leave it to them to progress to atomic resolution. Ramakrishnan writes that the ribosome was different because of its importance, and the fact that years had gone by "without much apparent progress towards an actual structure". By 1995, four groups around the world were in competition, and several others were contributing key techniques to improve crystallization or locate atoms.

Gene Machine: The Race to Decipher the Secrets of the Ribosome
VENKI RAMAKRISHNAN
Oneworld (2018)

In 1999, Ramakrishnan moved to the mecca of structural biology, the Medical Research Council Laboratory of Molecular Biology in Cambridge, UK. Scientists at this one lab had already garnered seven Nobel prizes. The institution has now racked up 11 — more than some scientifically advanced nations. Ramakrishnan focused on the smaller of the ribosome's subunits, which decodes mRNA before the larger subunit assembles the protein.

By 2000, he and his colleagues had solved its structure, finishing neck and neck with the team of Thomas Steitz at Yale, which had solved the large subunit. Previously seen as a rank outsider, Ramakrishnan was propelled into the limelight. A Nobel for the ribosome work was being openly discussed. There were up to six potential contenders for a prize that can be split only three ways. James Watson twice told Ramakrishnan that he shouldn't mind not being one of the three when the time came. Yet, in 2009, the time did come. Yonath, Ramakrishnan and Steitz shared that year's Nobel Prize in Chemistry.

Ramakrishnan credits his wife, the artist Vera Rosenberry, with keeping him grounded: on hearing of his prize, she said, "I thought you had to be really smart to win one of those!" He reflects on the disproportionate status such recognition brings. Suddenly, you're showered with other honours, and expected to pronounce sagely on everything from climate change to human cloning.

Some readers might take issue with how events or personalities are presented in *Gene Machine*. Yonath's pioneering work is fully acknowledged, for example. Yet, as Ramakrishnan's principal competitor, she sometimes appears in an unfavourable light. This is not an objective history of the field, but a highly personal account. As such, anyone who wants to know how modern science really works should read it. It's all here: the ambition, jealousy and factionalism — as well as the heroic late nights, crippling anxiety and disastrous mistakes — that underlie the apparently serene and objective surface represented by the published record. ■

Georgina Ferry *is the author of many books, including* Dorothy Hodgkin: A Life.
e-mail: georgina.ferry@gmail.com

Correspondence

Political solutions can beat algorithms

Regarding the computing arms race in US voter redistricting (W. K. Tam Cho *Nature* **558**, 487; 2018) between voting-rights advocates and users of sophisticated software for gerrymandering, a political solution could be simpler and more effective than a technological one.

Many criteria for electoral mapping compete with one another — such as population equality, compactness, maintenance of political and geographical boundaries and respect for communities of interest. Politicians can therefore argue for personally advantageous computer-optimized electoral maps while plausibly denying any nefarious intent to disenfranchise specific voters. However, turning the process over to an algorithm merely shifts the debate to the fairness of the algorithm itself. Computers might be impervious to the lure of power; their users are not.

Technology cannot readily resolve social problems that are based on conflicts over values and interests. To improve the ailing US political system, the country should instead consider a move to proportional representation — used in some form by many democratic nations. This would be much less susceptible to gerrymandering than the current winner-takes-all US voting system.

Daniel J. Rozell *Stony Brook University, New York, USA.*
daniel.rozell@stonybrook.edu

Shark's DNA should calm the waters

Alarm quickly spread after two children were bitten in July while swimming off Fire Island, New York, because great white sharks (*Carcharodon carcharias*) frequent the region. We can vouch from DNA analysis that another, relatively harmless, shark species was responsible for biting one of the individuals, and, in our view, probably the other as well.

We extracted DNA from a decontaminated fragment of shark tooth recovered from one of the bite wounds. Comparison with mitochondrial DNA sequences of some 900 species of cartilaginous fish enabled us to identify the DNA source as a sand tiger shark (*Carcharias taurus*; unpublished results). This shark is generally not considered to be dangerous to humans (J. I. Castro *The Sharks of North America* Oxford Univ. Press, 2011), despite its size (up to 3 metres long and weighing more than 200 kilograms).

The incidents occurred some 7 kilometres apart and within minutes of each other. This is not as surprising as it might seem — sand tiger sharks prey on schooling fishes, tracking them as they move inshore. This makes it more likely that the sharks will mistake nearby swimmers for prey and bite them. However, such random events are extremely rare.

Gavin J. P. Naylor* *University of Florida, Gainesville, USA.*
gnaylor@flmnh.ufl.edu
**On behalf of 4 correspondents (see go.nature.com/2nwptfh for full list).*

Screen evidence for power and bias

In our view, the four principles for making evidence synthesis more useful for policy would be strengthened by taking power and bias into account (C. A. Donnelly *et al. Nature* **558**, 361–364; 2018). Otherwise, the principles could fall short for issues that involve uncertain facts, disputed values, high stakes and urgent decisions — as in global biodiversity loss and climate change, for example.

Sometimes, complexities in scientific evidence allow several contrasting but equally valid interpretations. In such cases, there is a risk that privileged stakeholders associated with one way of thinking might unduly influence the particular values and interests prioritized in that synthesis.

Scientific aspirations, integrity and practices are crucial for challenging this authority. But if scientific disciplines and organizations deny or become complacent about their own forms of bias, then claims that purport to be definitive and objective could distort decision-making.

Evidence synthesis therefore needs to highlight contrasting valid framings of the best available evidence. A plural and conditional picture that is rigorous in embracing both social and natural sciences is more robust than single, evidence-based prescriptions. Analyses are inevitably influenced by politics. By improving transparency, those who hold power and privilege in and around science become more accountable.

We therefore suggest adding a fifth principle of open-mindedness, with mandates to examine the evidence from outside as well as inside science; to explain how contrasting values and interests yield divergent interpretations and prescriptions; and to evaluate the effects of power and privilege within established practices of evidence synthesis.

Andy Stirling *University of Sussex, Brighton, UK.*
Clive Mitchell *Scottish Natural Heritage, Battleby, Perth, UK.*
clive.mitchell@nature.scot

EU politicians must trust plant science

The latest ruling by the European Court of Justice requires that crops created using gene-editing techniques such as CRISPR must go through the same lengthy approval process as conventional genetically modified (GM) plants (see *Nature* **560**, 16; 2018). This has surprised many scientists, who are concerned that it will complicate promising applications of gene editing.

The court took existing legislation into account in arriving at its decision, but the situation has changed greatly since the first directives on GM organisms in 1990. Hundreds of millions of hectares have been planted worldwide with GM crops, providing extensive experience with such products. And techniques developed since could potentially solve important questions in biology and agriculture.

The court concluded that the European legislation considers the use of recombinant-DNA techniques in gene editing as sufficient grounds for classifying genome-edited plants as genetically modified. This could result in a costly approval process and might generate problems with unregulated genome-edited products imported from countries such as the United States.

One possibility would be to alter the legislation, but this could be difficult given current European politics. Another would be to revisit the European directives issued since 1990, which were based on a case-by-case scientific analysis of GM plants.

As members of the European Food Safety Authority's panel on GM organisms since its inception, we have witnessed a mounting distrust of scientific assessments. That has manifested with the approval of rules that demand a rigid analysis of GM plants. We need to reverse this trend, for example by acknowledging that approval of genome-edited plants calls for much less data than classic GM organisms, and by commanding greater respect for the work of scientific panels. This would promote scientifically sound risk analysis while complying with existing directives.

Josep M. Casacuberta, Pere Puigdomènech *Centre for Research in Agricultural Genomics, Barcelona, Spain.*
pere.puigdomenech@cragenomica.es

PLANETARY SCIENCE

Jupiter's magnetic field revealed

The magnetic field of Jupiter has been found to be different from all other known planetary magnetic fields. This result could have major implications for our understanding of the interiors of giant planets. SEE LETTER P.76

CHRIS JONES

NASA's Juno spacecraft is currently mapping Jupiter's magnetic field in unprecedented detail. Because the field originates in the planet's interior, it can provide insights into what is going on beneath the spectacular swirling clouds in the planet's surface layers. On page 76, Moore *et al.*[1] analyse data from Juno and find that Jupiter's magnetic field is substantially different in the planet's northern and southern hemispheres. The authors consider what might be happening in the planet's interior to account for this asymmetry.

Juno reached Jupiter on 4 July 2016, and it has been gathering data that are transforming our understanding of the planet's deep interior. Previously, we had only a broad-brush overview of Jupiter's magnetic field[2]. Juno has brought the picture into much sharper focus, allowing a revised model of the field to be constructed[3]. These advances were possible owing to the close approach that Juno makes to Jupiter — the spacecraft flies only about 4,000 kilometres above Jupiter's surface as it dives into the planet's gravitational field once every 53 days[4].

Jupiter has the strongest planetary magnetic field in the Solar System. Ironically, this field is the biggest threat to the Juno mission. High-energy particles from the Sun are trapped in the field, producing a hazard that is dangerous to the electronics on which the mission depends. Fortunately, Juno was designed with protection against this and has survived so far.

The magnetic field of Jupiter is maintained by electric currents that flow in the planet's interior. Jupiter is made up mainly of hydrogen and helium, so it is quite surprising that it can conduct electricity at all. However, the extremely high pressure and density in the planet enable hydrogen to enter a state known as metallic hydrogen[5]. Metallic hydrogen has an electrical conductivity similar to that of metals, allowing electric currents to flow.

Giant planets take billions of years to cool down after they are formed. Consequently, there is as much heat coming out of Jupiter's interior as is received by the planet from the Sun. This heat is carried by convection currents, which stir the interior and produce the swirling clouds and storms — such as the Great Red Spot — that are so beautifully captured by Juno's cameras. The convection-driven flows of fluid in the interior are slower than the surface winds, but they are strong enough to generate Jupiter's magnetic field by a process called dynamo action[6,7].

Earth's magnetic field is also produced by convection-driven flows in the planet's interior, but it is the planet's liquid-iron core that allows electric currents to flow. The fields of both Jupiter and Earth are mainly dipolar — the radial component of the field is mostly positive in the northern hemisphere and mostly negative in the southern hemisphere, as if the planet contained a bar magnet (Fig. 1a). Moore and colleagues report that the non-dipolar part of Jupiter's field is confined almost entirely to the northern hemisphere (Fig. 1b). This is in stark contrast to Earth's field, for which the non-dipolar part is evenly distributed between the two hemispheres.

Moore *et al.* suggest several possible explanations for the morphology of Jupiter's magnetic field. One explanation concerns Jupiter's core, the nature of which is still a mystery. Some models of the planet assume a compact core with a mass about five times that of Earth[8]. But a much larger, dilute core is also feasible[9], and could affect field generation.

Another explanation is that there are one or more stable layers of fluid deep inside Jupiter. Saturn is thought to have a stable layer in its interior, which could account for why its magnetic field is almost completely symmetrical about the planet's rotation axis[10] — vastly different from the fields of Jupiter and Earth. In Jupiter, these stable layers might be regions in which the composition of the fluid changes, partitioning the planet's interior into zones. If the transition regions contained a helium concentration gradient, they could be bottom heavy, altering the fluid flow inside the planet and therefore the magnetic field.

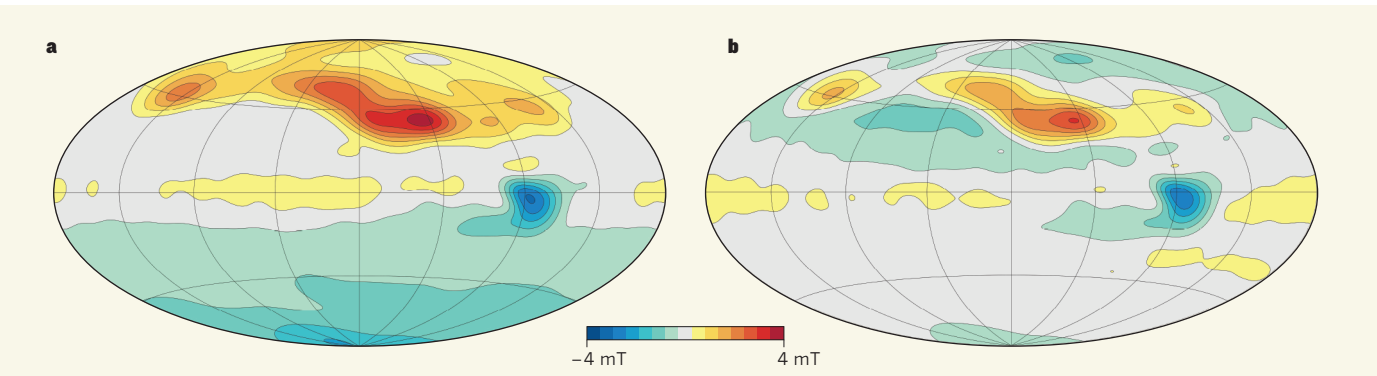


Figure 1 | Maps of Jupiter's magnetic field. **a**, In the northern hemisphere of Jupiter, the radial component of the planet's magnetic field points mainly in the positive (outwards) direction (yellow–red shades). Conversely, in the southern hemisphere, the radial component points predominantly in the negative (inwards) direction (green–blue shades). Such a configuration is known as a dipole. The colour scale depicts the strength of the radial magnetic field in units of millitesla. **b**, Moore *et al.*[1] report that the non-dipolar part of Jupiter's radial magnetic field is almost entirely concentrated in the northern hemisphere — unlike all other known planetary magnetic fields. The maps in **a** and **b** illustrate the magnetic field at a distance of 90% of Jupiter's radius from the planet's centre, under the assumption that substantial electric currents in the planet all reside at distances closer to the centre. (Adapted from Fig. 1e and Fig. 3a of ref. 1.)

To investigate how planetary magnetic fields are generated, it is now possible to solve the fundamental equations that govern the fluid flows and the magnetic fields inside planets. The basic principles of dynamo action were laid down a century ago[11], but solving the fluid–dynamo equations proved difficult. Computers have been able to handle the calculations required to model Earth's dynamo only since 1995 (ref. 12). Nevertheless, much progress has been made, and computational models of dynamos can now capture many of the characteristics of Earth's magnetic field[13].

In the past five years, these models have been adapted to deal with the large variations in density between the interior and atmosphere of Jupiter[6,7], and can now be compared with the field inferred by Moore and colleagues. However, dynamo models depend on the internal structure of the planet, which in turn depends on the planet's thermodynamic properties, electrical-conductivity profile and composition. Although these issues have been extensively explored, some uncertainty remains. Models of fields that are dipolar but broadly symmetric about the equator have been developed[6], as have models of fields that are asymmetric but not dipolar[14]. The challenge is therefore to formulate models of fields that are both asymmetric and dipolar.

Moore and colleagues' suggested explanations for Jupiter's field morphology can now be tested by dynamo modellers to discover whether the explanations are indeed compatible with Juno's observations. Exciting times lie ahead for the study of the interiors of giant planets, as modellers digest the information coming from Juno and begin to work out a clearer picture of the inside of Jupiter. ■

Chris Jones *is in the Department of Applied Mathematics, University of Leeds, Leeds LS2 9JT, UK.*
e-mail: cajones@maths.leeds.ac.uk

1. Moore, K. M. *et al. Nature* **561**, 76–78 (2018).
2. Ridley, V. A. & Holme, R. *J. Geophys. Res. Planets* **121**, 309–337 (2016).
3. Connerney, J. E. P. *et al. Geophys. Res. Lett.* **45**, 2590–2596 (2018).
4. Bolton, S., Levin, S. & Bagenal, F. *Geophys. Res. Lett.* **44**, 7663–7667 (2017).
5. Wigner, E. & Huntington, H. B. *J. Chem. Phys.* **3**, 764–770 (1935).
6. Jones, C. A. *Icarus* **241**, 148–159 (2014).
7. Gastine, T., Wicht, J., Duarte, L. D. V., Heimpel, M. & Becker, A. *Geophys. Res. Lett.* **41**, 5410–5419 (2014).
8. French, M. *et al. Astrophys. J. Suppl. Ser.* **202**, 5–16 (2012).
9. Wahl, S. M. *et al. Geophys. Res. Lett.* **44**, 4649–4659 (2017).
10. Stevenson, D. J. *Geophys. Astrophys. Fluid. Dyn.* **21**, 113–127 (1982).
11. Larmor, J. *Rep. Br. Assoc. Adv. Sci.* **87**, 159–160 (1919).
12. Glatzmaier, G. A. & Roberts, P. H. *Nature* **377**, 203–209 (1995).
13. Schaeffer, N., Jault, D., Nataf, H.-C. & Fournier, A. *Geophys. J. Int.* **211**, 1–29 (2017).
14. Dietrich, W. & Jones, C. A. *Icarus* **305**, 15–32 (2018).

IMMUNOLOGY

An immune response with a sweet tooth

A previously unknown pathway that enables mammalian cells to recognize infection and trigger an immune response requires a kinase enzyme in the host cell to bind a sugar molecule produced by infecting bacteria. SEE LETTER P.122

JOHN-DEMIAN SAUER

Bacterial infections are a major cause of disease and death worldwide. The innate branch of the mammalian immune system, which recognizes and reacts to general characteristics of pathogenic organisms, has a key protective role. On page 122, Zhou *et al.*[1] describe a mechanism by which the innate immune system is activated in response to bacterial sugar molecules. This finding broadens our understanding of the types of molecule that can be recognized as hallmarks of bacterial infection and the host proteins that can recognize such molecules.

A key advance in our understanding of how the innate immune system functions was the identification of proteins called pattern-recognition receptors (PRRs), which recognize 'non-self' molecules termed pathogen-associated molecular patterns (PAMPs). Beginning with the Toll and Toll-like receptor PRRs[2–4] in the late 1990s, the identification of PRRs and the PAMPs that they recognize has proceeded at a breathtaking pace.

A key function of PRRs is to help drive the expression of secreted proteins called cytokines, which alert the immune system to the presence of infection. The transcription factor NF-κB is a central regulator of cytokine expression. Zhou and colleagues studied human cells grown *in vitro* to try to identify pathways that activate NF-κB in response to infection by the bacterium *Yersinia pseudotuberculosis*. This bacterium has a needle-like, multiprotein structure called a type III secretion system (T3SS), which is required for the direct transfer of bacterial proteins into host cells. T3SSs are evolutionarily conserved in many pathogenic bacteria.

Zhou *et al.* took an unbiased approach and screened a collection of *Y. pseudotuberculosis* genetic mutants to identify bacterial genes that are linked to NF-κB activation in response to infection. This led the authors to focus on the enzyme HldE, which catalyses steps in the biosynthetic pathway that generates lipopolysaccharide (LPS) molecules.

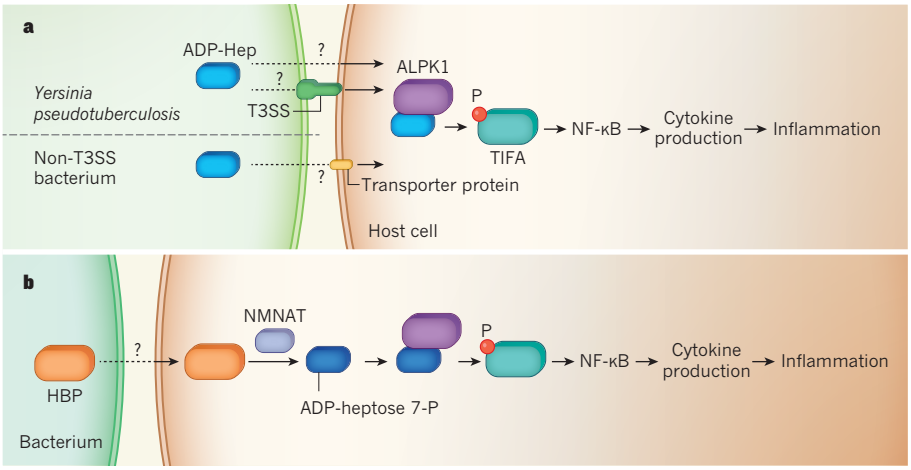


Figure 1 | Bacterial sugars trigger a host immune response. **a**, Zhou *et al.*[1] demonstrate that in bacteria such as *Yersinia pseudotuberculosis*, which has a multiprotein complex called a type III secretion system (T3SS), and in other bacterial species lacking a T3SS, the sugar molecule ADP-β-D-manno-heptose (ADP-Hep) can enter a host cell, by an unknown route (possibly through a transporter protein), and can trigger a signalling pathway that drives inflammation. When ADP-Hep enters the host cell, it binds to ALPK1, which activates the protein TIFA by adding a phosphate group (P) to it. The downstream signalling pathway, not all the steps of which are shown, leads to activation of the protein NF-κB, which drives the expression of cytokine proteins that stimulate an immune response to the infection. **b**, The authors also report that if the bacterially produced sugar D-glycero-β-D-manno-heptose 1,7-bisphosphate (HBP) enters the host cell (by a route that remains to be determined), it can be converted by host enzymes of the NMNAT family into the molecule ADP-heptose 7-P. This binds to ALPK1 and activates the same pathway as that activated by ADP-Hep.

LPS is an essential component of the cell surface of a subset of bacterial pathogens called Gram-negative bacteria.

Using genetically mutated bacteria and purified sugar molecules, the authors sought to pinpoint the molecules in the LPS biosynthetic pathway that stimulate NF-κB activation. They found that the presence of bacterial sugars, including ADP-β-D-manno-heptose (ADP-Hep) and D-glycero-β-D-manno-heptose 1,7-bisphosphate (HBP), in the host-cell cytoplasm triggered NF-κB activation. This is consistent with a study[5] of *Neisseria meningitidis* bacteria that demonstrated that HBP can trigger NF-κB responses in host cells. Crucially, Zhou *et al*. showed that ADP-Hep is 100 times more potent than is HBP at activating NF-κB. They found that addition of ADP-Hep to the extracellular environment of host cells can activate NF-κB, suggesting that dedicated host-cell transporter proteins deliver ADP-Hep to the host's cytoplasm.

No PRR was known to recognize ADP-Hep. To search for one, the authors used a gene-editing approach to conduct a screen in which they generated random mutations in host cells and tested whether the mutations affected ADP-Hep recognition. They uncovered two candidate genes that respectively encode the kinase enzyme ALPK1 and the protein TIFA, and showed that these are required for NF-κB activation in response to ADP-Hep in host cells (Fig. 1). A previous study had revealed[5] that TIFA is required for recognition of HBP from *N. meningitidis*. ALPK1 and TIFA signalling has also been linked to HBP-dependent host activation of NF-κB in response to infection by the bacteria *Shigella flexneri*[6] and *Helicobacter pylori*[7]. Using biochemical approaches, Zhou and colleagues demonstrated that ADP-Hep binds directly to the amino terminus of ALPK1. The authors solved the X-ray crystal structure of ALPK1 in a complex with ADP-Hep, and validated their structural model by testing the effect of mutations in ALPK1 that were predicted to impair its binding to ADP-Hep.

Zhou *et al*. also generated ALPK1-deficient mice. The NF-κB-dependent production of cytokines was significantly reduced in these animals after challenge with either ADP-Hep or the pathogenic bacterium *Burkholderia cenocepacia*, compared with results seen in animals that were not deficient in ALPK1. Moreover, the number of bacteria in the lungs of mice infected with *B. cenocepacia* was higher in ALPK1-deficient animals than in wild-type mice.

Perhaps Zhou and colleagues' most striking finding is that mammalian adenylyltransferase enzymes, specifically those of the NMNAT family, catalyse a reaction that converts HBP into a molecule called ADP-heptose 7-P, which can act as a ligand by binding to ALPK1. Previous work[5] had suggested that HBP is a PAMP that can directly activate NF-κB. Although HBP can be defined as a PAMP, given that it is a bacterially derived molecule that triggers a host response, Zhou and colleagues' data indicate that HBP must be converted to ADP-heptose 7-P by host enzymes to trigger this response. The authors report slight differences in the way in which ADP-Hep and ADP-heptose 7-P bind to ALPK1, and use these differences to demonstrate why ADP-Hep and not HBP or ADP-heptose 7-P is the relevant ligand for ALPK1-mediated NF-κB activation, at least in *Y. pseudotuberculosis* infection.

Zhou and colleagues' findings have important implications. Evidence that ADP-Hep is a PAMP adds to a growing awareness that bacterial metabolites can act as PAMPs. Given that ADP-Hep is needed to synthesize an essential component of the outer membrane of most Gram-negative bacteria, this makes it an ideal PAMP. However, it is not known how this molecule, which is normally found inside the bacterium, reaches the cytoplasm of the host cell. In *Y. pseudotuberculosis*, this process requires the T3SS, although it is unclear whether ADP-Hep is actively transported or accidentally leaks through the T3SS, or whether it enters by the pores that the T3SS generates in the host-cell membrane.

The authors report that bacterial species that lack a T3SS can still trigger the ALPK1 pathway in an ADP-Hep-dependent manner, consistent with the ability of purified ADP-Hep to activate the pathway by an extracellular route. This suggests that a dedicated transport system might exist that allows the host cell to sample its extracellular surroundings for the presence of this PAMP, similar to the way in which certain extracellular PAMPs are transported to the cytoplasm for recognition by host proteins[8].

Why does bacterial ADP-Hep exposure occur if it activates the innate immune system? Perhaps its release is needed to fulfil some as yet unknown function. Pathogens often evolve mechanisms to evade or thwart an immune-system response. If pathogens have evolved strategies to avoid triggering an ADP-Hep-mediated immune response, understanding such strategies might suggest new therapeutic approaches to fight bacterial infections.

The authors' observation that host enzymes can convert bacterial metabolites that have poor immune-activating characteristics into potent PAMPs offers a new perspective on the evolutionary battle between pathogens and their hosts. Although Zhou *et al*. show that ADP-Hep is the relevant immune-triggering ligand for *Y. pseudotuberculosis* infections, it remains to be seen whether HBP is converted into ADP-heptose 7-P during other bacterial infections. This issue is particularly relevant for pathogens (for example, *Shigella*) that invade the host-cell cytoplasm and that might shed PAMPs such as HBP directly into the cytoplasm. Zhou and colleagues' work also offers a fresh perspective on the types of molecule that can act as PAMPs or their PRRs, and where and how researchers should be searching for such molecules. ■

John-Demian Sauer *is in the Department of Medical Microbiology and Immunology, University of Wisconsin–Madison, Madison, Wisconsin 53706, USA.*
e-mail: sauer3@wisc.edu

1. Zhou, P. *et al. Nature* **561**, 122–126 (2018).
2. Lemaitre, B., Nicolas, E., Michaut, L., Reichhart, J.-M. & Hoffmann, J. A. *Cell* **86**, 973–983 (1996).
3. Medzhitov, R., Preston-Hurlburt, P. & Janeway, C. A. Jr *Nature* **388**, 394–397 (1997).
4. Poltorak, A. *et al. Science* **282**, 2085–2088 (1998).
5. Gaudet, R. G. *et al. Science* **348**, 1251–1255 (2015).
6. Milivojevic, M. *et al. PLoS Pathog.* **13**, e1006224 (2017).
7. Zimmermann, S. *et al. Cell Rep.* **20**, 2384–2395 (2017).
8. Lee, J. *et al. J. Biol. Chem.* **284**, 23818–23829 (2009).

This article was published online on 15 August 2018.

CHEMICAL BIOLOGY

DNA tags light up sugars on proteins

Methods for imaging sugars attached to proteins — the protein glycoforms — are of interest because glycoforms affect protein movement and localization in cells. A versatile approach is now reported that uses DNA as molecular identity tags.

TADASHI SUZUKI

The attachment of sugar molecules to proteins is one of the most common protein modifications, found in all domains of life. Sugars attached to proteins are called glycans, and modulate the physicochemical and physiological properties of the carrier proteins[1]. But tracking and visualizing glycoforms — the specific patterns of sugars attached to a protein — in cells is challenging, particularly if you want to visualize several different glycoforms at once. Writing in *Angewandte Chemie*, Li *et al*.[2] now report a method for doing this that relies on the dynamic interactions of a set of DNA codes.

Since the early 1990s, the use of fluorescent tags as labels for proteins has revolutionized

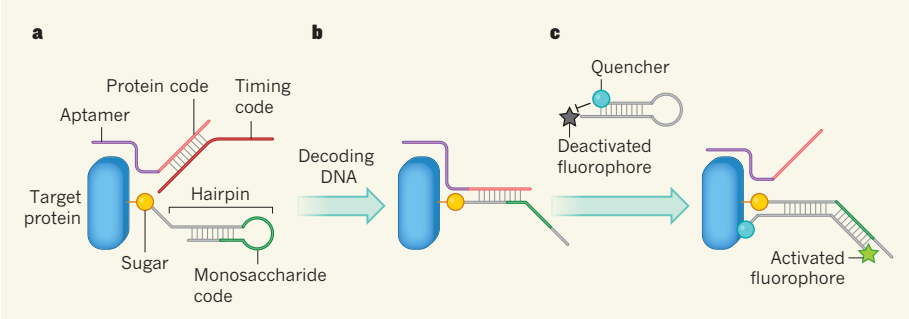


Figure 1 | A method for visualizing sugars on proteins[2]. **a**, A DNA sequence (the protein code) is bound to the target protein through a sequence called an aptamer. The protein code is hybridized (forms a double helix) with a second sequence, called the timing code. A sugar attached to the protein is covalently attached to a 'hairpin' DNA, which contains a masked sequence that identifies the sugar (the monosaccharide code). **b**, A 'decoding' DNA molecule is added that hybridizes with the timing code (not shown), releasing the protein code so that it hybridizes with part of the hairpin. The hairpin opens, unmasking the monosaccharide code. **c**, A second hairpin DNA is added, which is complementary to the monosaccharide code; it also bears a fluorescent molecule (a fluorophore) at one end and a quencher molecule at the other, which deactivates fluorescence. The hairpin hybridizes with the DNA containing the monosaccharide code, opening the hairpin and allowing the fluorophore to fluoresce. The protein code is simultaneously exposed by this process, and can take part in another cycle of reactions.

how cell biologists analyse protein movement and localization in cells[3,4]. But even though the types of glycan attached to proteins can affect their movement and localization, it has been difficult to visualize any particular glycoform. One way in which researchers have attempted to solve this problem is by using a technique called fluorescence resonance energy transfer (FRET). In this technique, a fluorescent molecule (a fluorophore) is attached to a protein of interest and a second fluorophore is attached to a specific sugar; fluorescence occurs only if the two molecules come into close proximity through the attachment of the sugar to the protein[5–8]. However, the need to use two different fluorophores can limit applications, for example by making it difficult to detect multiple glycoforms of a protein in the same experiment.

Li *et al.* overcome this problem using an approach that they describe as a hierarchical coding strategy, in which multiple single-stranded DNA molecules are used as identification codes to visualize specific sugars attached to a chosen protein (Fig. 1). The first DNA molecule used in the authors' system contains a sequence (known as an aptamer) that specifically binds to the target protein. The aptamer is attached to another sequence (the protein code) that identifies the protein. A second DNA molecule, called the timing code, contains a sequence that is complementary to the protein code, and that therefore hybridizes (forms a double helix) with it.

The third DNA molecule used in Li and colleagues' system contains three segments. The first segment is complementary to the protein code. This is attached to a second sequence called the monosaccharide code, which identifies a specific sugar. The third segment has a sequence that enables the complete strand to form a structure known as a hairpin, which masks the monosaccharide code. The hairpin DNA is covalently attached to the sugar identified by the monosaccharide code. If the hairpin-bearing sugar is in turn attached to the target protein, this can bring the hairpin into close proximity with the double helix formed by the protein and timing codes.

The final key component of Li and colleagues' system is another hairpin DNA, which contains a complementary sequence to the monosaccharide code and a sequence that can displace the protein code from a double helix. The hairpin also has a fluorophore attached at the 5′ end, and a 'quencher' molecule at the 3′ end. The quencher stops the fluorophore from fluorescing when the hairpin is closed, but allows fluorescence when the hairpin opens.

So how do all these components interact to decode the crucial DNA identifiers and allow glycoforms to be visualized? The process is triggered when a single-stranded DNA that is complementary to the timing code is added to the system. This DNA hybridizes with the timing code, thus displacing and exposing the protein code. The exposed protein code then hybridizes with the complementary sequence in the hairpin attached to the sugar, opening up the hairpin and unmasking the monosaccharide code.

When the fluorophore-carrying hairpin is added to the system, the unmasked monosaccharide code hybridizes with the complementary DNA sequence in that hairpin. The hairpin therefore opens up, allowing its fluorophore to fluoresce: in effect, a fluorescent tag has been attached to the sugar, allowing it to be detected. The hybridization also unmasks the protein code, making it available for another reaction cycle. The key element of Li and colleagues' system is that the protein code is physically associated with the target protein, because this ensures that only hairpin-bearing sugars that are attached or close to the protein can become fluorescent.

The authors confirmed that the chain of reactions occurs in a cell-free system *in vitro* and used it to identify two glycoforms of the MUC1 protein: MUC1 decorated with the sugar fucose, or with another sugar called sialic acid. Crucially, the authors also showed that the fluorescent signals can be generated and detected on cells that had been modified using a method known as metabolic labelling[9] to incorporate hairpin-bearing sugars.

An advantage of this method is that, because the choice of DNA sequences that can be used as labels is effectively infinite, many different glycoforms can be imaged, as long as the proteins and sugars can be specifically labelled with their own DNA codes. Moreover, the authors clearly showed that sialylated and fucosylated MUC1 could be simultaneously detected using their method. One potential limitation, however, is that the DNA used was not observed to be transported into cells through natural processes, suggesting that intracellular glycoforms cannot be detected by this method. This could actually be an advantage for studies that focus on cell-surface proteins.

A few issues will need to be clarified in future studies. For example, the efficiency of the decoding process is unclear. It is also not known whether sugars on molecules next to the target proteins might sometimes become fluorescent, as a result of DNA hybridization between the protein code and hairpins attached to sugars on neighbouring proteins. Because a large number of glycans are attached to MUC1, the method might not need to be highly efficient to generate a detectable fluorescent signal for this protein, and any minute signals produced from neighbouring molecules would not be a serious problem. However, further experiments using other glycoproteins that have fewer sugars attached are needed to validate the method fully.

Given that both the above issues might depend largely on the length of the DNA chains used, careful design of the DNA codes and of the aptamers will be essential for ensuring the specific detection of other glycoforms. The practical advantages and disadvantages of the new technique compared with other strategies for glycoform imaging that have been reported in the past few years — including two methods reported by workers from the same group as Li *et al.*[10,11] — also remain to be explored.

Nevertheless, Li and colleagues' hierarchical coding strategy for glycoform imaging shows great potential, and could be an important step in the development of a system analogous to the use of green fluorescent proteins for protein tagging — which is now standard practice for biologists. The ultimate goal is to visualize glycoforms in a way that will enable us to see what we want to see, rather than only what can be seen. ■

Tadashi Suzuki *is in the Glycometabolic Biochemistry Laboratory, RIKEN Cluster*

for Pioneering Research, Saitama 351-0198, Japan.
e-mail: tsuzuki_gm@riken.jp

1. Varki, A. *Glycobiology* **27**, 3–49 (2017).
2. Li, S. *et al. Angew. Chem. Int. Edn* https://doi.org/10.1002/anie.201807054 (2018).
3. Miyawaki, A. *Nature Rev. Mol. Cell Biol.* **12**, 656–668 (2011).
4. Hori, Y. & Kikuchi, K. *Curr. Opin. Chem. Biol.* **17**, 644–650 (2013).
5. Haga, Y. *et al. Nature Commun.* **3**, 907 (2012).
6. Lin, W., Du, Y., Zhu, Y. & Chen, X. *J. Am. Chem. Soc.* **136**, 679–687 (2014).
7. Belardi, B. *et al. Angew. Chem. Int. Edn* **52**, 14045–14049 (2013).
8. Doll, F. *et al. Angew. Chem. Int. Edn* **55**, 2262–2266 (2016).
9. Laughlin, S. T. & Bertozzi, C. R. *Nature Protocols* **2**, 2930–2944 (2007).
10. Wu, N., Bao, L., Ding, L & Ju, H. *Angew. Chem. Int. Edn* **55**, 5220–5224 (2016).
11. Hui, J. *et al. Angew. Chem. Int. Edn* **56**, 8139–8143 (2017).

ECOLOGY

A systemic problem with pesticides

Exposure to a sulfoximine-based pesticide has substantial adverse effects on bumblebee colonies. This finding suggests that concerns over the risks of exposing bees to insecticides should not be limited to neonicotinoids. SEE LETTER P.109

NIGEL E. RAINE

Agricultural intensification has increased our reliance on pesticides, including insecticides. Although insecticides are useful for controlling crop damage caused by insect pests, they can also affect beneficial insects, potentially impairing their ability to control pests and pollinate crops[1] — qualities on which farmers rely. Indeed, increases in insecticide use are one of several major factors implicated in the worldwide declines of insect pollinators[2]. A commonly used class of insecticide called neonicotinoids has hit the headlines because of its impacts on bees. Siviter *et al.*[3] report on page 109 that a potential neonicotinoid replacement, the sulfoximine-based insecticide sulfoxaflor, also harms these crucial pollinators.

Insect pollinators that forage on neonicotinoid-treated plants can be exposed to small amounts of insecticide each time they or their larvae feed on pollen and nectar[4,5]. Although such chronic neonicotinoid exposure typically does not kill bees, it can have sublethal effects — impairing a range of behaviours such as learning and foraging[4–8], affecting nesting success, colony development and reproduction[7–12], and reducing pollination levels[13]. Because of this, substantial restrictions on neonicotinoid use have been introduced in some regions of the world, particularly Europe. Such restrictions might seem to be good news for bee health — but only if the insecticides that replace neonicotinoids are less harmful to insect pollinators.

Similar to neonicotinoids, sulfoximine-based insecticides are absorbed and systemically distributed throughout the plant. Sulfoxamines are one candidate to replace neonicotinoids[14], and have already been widely approved for use. Siviter and colleagues set out to assess the sublethal effects of sulfoxaflor on the agriculturally important pollinator *Bombus terrestris*. This bumblebee is common in the wild, and is also reared commercially for crop pollination. Although it is convenient to use commercially reared colonies for experiments, the authors chose to use wild colonies — a decision that should be lauded because it enhances the ecological realism of their study.

Siviter *et al.* collected 332 wild queen bumblebees, assessed them for parasites and used 249 uninfected individuals to start colonies in the laboratory. The authors succeeded in rearing colonies from 52 queens, providing a robust sample size for their experiment. They then randomly allocated pairs of size-matched bee colonies to either control or insecticide-exposure groups. The colonies fed at will for two weeks on either sugar water alone or sugar water containing five parts per billion of sulfoxaflor (a concentration found in the nectar of crops sprayed with sulfoxaflor), before being moved outdoors, so that the researchers could monitor bee behaviour and colony development under field conditions.

The team found that sulfoxaflor exposure had substantial and consistent effects on the rate of colony growth, which became apparent after just two to three weeks in the field. Sulfoxaflor-exposed colonies produced fewer female workers than did control colonies. They also produced 54% fewer reproductive offspring. This substantial difference was predominantly driven by a decrease in the total number of males produced, but also reflects the fact that all of the 36 new queens produced came from just 3 of the control colonies. Such strong variation in queen production among control colonies is not unexpected, but the lack of queen production by any of the insecticide-exposed colonies is concerning, because queens are needed to start new colonies in the following year.

These impairments in colony growth and reproduction are similar to those observed in comparable neonicotinoid-exposure studies[8–10, 12,15,16]. This similarity might be expected, given that both insecticide classes affect insects by binding to the same neurotransmitter receptors[14]. But whereas the

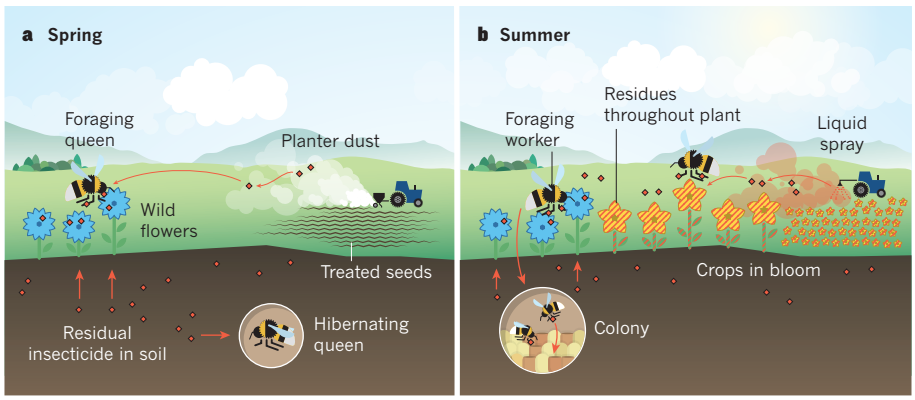


Figure 1 | Routes of bumblebee exposure to insecticides. Siviter *et al.*[3] have investigated how exposure to the insecticide sulfoxaflor affects bumblebee colonies, using a combined laboratory–field protocol. There are multiple potential routes of exposure to systemic insecticides. **a**, In spring, insecticide-treated seeds are sown. Contaminated dust from seed planters drifts across fields, and lands on wild flowers (insecticide residues are indicated by red diamonds, routes of spread by red arrows). Residual insecticide in the soil from the previous year might affect queen bumblebees hibernating in the soil, or be taken up by wild flowers, leading to exposure of foraging queens that consume contaminated nectar and pollen. **b**, In summer, crops grown from treated seeds bloom, producing contaminated nectar and pollen (red stripes). Spray treatments can increase insecticide levels on crops and on nearby wild flowers. Foraging worker bees ingest insecticide-laced nectar and pollen from both treated crops and contaminated wild flowers[17,18], and are exposed through contact with sprayed plant tissue when foraging on crops. Workers take insecticide-laced pollen and nectar back to the colony, where it is ingested by larvae (not shown).

effects on bumblebee colonies exposed to neonicotinoids seem to be driven by impaired pollen foraging[7,8] (leading to limited nutrition for larvae), the authors found no evidence that sulfoxaflor exposure caused significant differences in foraging performance. Perhaps early-stage colony growth and subsequent reproductive output were affected by sulfoxaflor toxicity to developing larvae, or by some other indirect mechanism — either way, the timing of declines in colony growth rate suggests that chronic sublethal stress at an early stage resulted in substantially reduced colony reproduction[15].

Correctly determining the effects of insecticides relies on accurate assessments of exposure, which varies depending on whether chemicals are applied by spray, soil drench or seed treatment (Fig. 1). For example, spray applications can lead to relatively high levels of exposure for a few days, whereas seed treatments can result in low-level, chronic exposure through residues in nectar and pollen[4,5]. The authors based exposure to sulfoxaflor in their experiment on a scenario in which bees ingest nectar from crop flowers following a spray application — currently, the most common mode of application for this insecticide class.

However, this scenario discounts any exposure from contact with plant tissues or dietary exposure from crop pollen, and assumes that bees forage only on sulfoxaflor-treated crops — all factors that could affect exposure levels. Moreover, exposure profiles would probably differ if sulfoxaflor were applied as a soil drench or seed treatment (an increasingly likely outcome following recent and probable future neonicotinoid regulation). Exposure could also be affected if sulfoxaflor, applied as a seed treatment or soil drench, moves outside crop fields and is absorbed by wild plants and contaminates their nectar and pollen, as reported for neonicotinoid seed treatments[17,18]. More data on sulfoxaflor concentrations in the nectar and pollen of bee-attractive crops are needed for an accurate assessment of the implications of sulfoxaflor use.

Nonetheless, Siviter *et al.* provide a valuable first step towards understanding the effects of sulfoxaflor exposure on bees. Future discussions must be broader than two-way comparisons of neonicotinoids and sulfoximines, because other classes of systemic insecticide (such as butenolides and anthranilic diamides) are also in agricultural use. It is vital to ascertain which of these insecticide classes represents the lowest potential risk to pollinators. A major part of the answer depends on how comparative risk assessments are undertaken, including which of the 20,000 living bee species are considered, because there is substantial variation in physiology, behaviour and ecology between these species. Such differences — particularly the extent to which species are social — might affect the bees' sensitivity to insecticides[10,12,19]. For instance, low-level insecticide exposure might have more impact on solitary bees than on highly social colonies that have an abundance of workers.

Finally, commercially reared pollinators (particularly honeybees) feature prominently in global agriculture, but cannot provide all of the crop-pollination services needed[20]. Wild pollinators, including bumblebees and solitary bees, have a crucial, undervalued role that is likely to become increasingly important as our crop-pollination demands rise[1,20]. Our understanding of the risks to pollinators, and the choices we make about pest control, must evolve to reflect and balance these realities. There are no risk-free choices, but with more information such as that provided by Siviter and colleagues, we can make the most appropriate decisions about how to produce the food we need without inflicting irreparable damage on the global environment and the essential ecosystem services (such as pollination) on which we depend. ■

Nigel E. Raine *is at the School of Environmental Sciences, University of Guelph, Guelph, Ontario N1G 2W1, Canada.*
e-mail: nraine@uoguelph.ca

1. Garibaldi, L. A. *et al.* Science **339**, 1608–1611 (2013).
2. Vanbergen, A. J. & the Insect Pollinators Initiative. *Front. Ecol. Environ.* **11**, 251–259 (2013).
3. Siviter, H., Brown, M. J. F. & Leadbeater, E. *Nature* **561**, 109–112 (2018).
4. Godfray, H. C. J. *et al. Proc. R. Soc. B* **281**, 20140558 (2014).
5. Godfray, H. C. J. *et al. Proc. R. Soc. B* **282**, 20151821 (2015).
6. Stanley, D. A. & Raine, N. E. *Funct. Ecol.* **30**, 1132–1139 (2016).
7. Gill, R. J., Ramos-Rodriguez, O. & Raine, N. E. *Nature* **491**, 105–108 (2012).
8. Stanley, D. A., Russell, A. L., Morrison, S. J., Rogers, C. & Raine, N. E. *J. Appl. Ecol.* **53**, 1440–1449 (2016).
9. Whitehorn, P. R., O'Connor, S., Wackers, F. L. & Goulson, D. *Science* **336**, 351–352 (2012).
10. Rundlöf, M. *et al. Nature* **521**, 77–80 (2015).
11. Baron, G. L., Jansen, V. A. A., Brown, M. J. F. & Raine, N. E. *Nature Ecol. Evol.* **1**, 1308–1316 (2017).
12. Woodcock, B. A. *et al. Science* **356**, 1393–1395 (2017).
13. Stanley, D. A. *et al. Nature* **528**, 548–550 (2015).
14. Brown, M. J. F. *et al. PeerJ* **4**, e2249 (2016).
15. Bryden, J., Gill, R. J., Mitton, R. A. A., Raine, N. E. & Jansen, V. A. A. *Ecol. Lett.* **16**, 1463–1469 (2013).
16. Ellis, C., Park, K. J., Whitehorn, P., David, A. & Goulson, D. *Environ. Sci. Technol.* **51**, 1727–1732 (2017).
17. Tsvetkov, N. *et al. Science* **356**, 1395–1397 (2017).
18. Nicholls, E. *et al. Environ. Sci. Technol.* https://doi.org/10.1021/acs.est.7b06573 (2018).
19. Arena, M. & Sgolastra, F. *Ecotoxicology* **23**, 324–334 (2014).
20. Aizen, M. A. & Harder L. D. *Curr. Biol.* **19**, 915–918 (2009).

This article was published online on 15 August 2018.

STRUCTURAL BIOLOGY

Spotlight on proteins that aid malaria

The multiprotein complex PTEX enables malaria-causing parasites to survive inside red blood cells. Studies reveal how PTEX assembles, and identify a function for one of the complex's proteins, EXP2. SEE ARTICLE P.70

TANIA F. DE KONING-WARD

Malaria is caused by the parasite *Plasmodium falciparum*. For part of its life cycle, this organism resides inside human red blood cells in a membrane-bound compartment called a vacuole. To survive, multiply and evade an immune response in this environment, *P. falciparum* must transport nutrients and proteins across the vacuolar membrane[1]. On page 70, Ho *et al.*[2] report the structure of the parasite PTEX complex, which resides on the vacuolar membrane and facilitates the export of proteins from the vacuole to the cytoplasm of red blood cells[3]. And in a paper in *Nature Microbiology*, Garten *et al.*[4] reveal that the protein EXP2, which forms part of the PTEX protein-conducting channel located in the vacuolar membrane, can also form a channel that facilitates nutrient transfer across the membrane. These insights into the structure and function of key proteins that aid the survival of *P. falciparum* might help efforts to develop new antimalarial drugs.

PTEX consists of five proteins[3]: HSP101, PTEX150, EXP2, PTEX88 and TRX2. Multiple HSP101, PTEX150 and EXP2 molecules assemble to form the core part of PTEX[3,5]. It has been predicted that HSP101 unfolds proteins destined for export, and provides the energy needed for cargo to pass through the vacuolar-membrane-spanning part of the channel, which is proposed[3,6] to consist of EXP2. PTEX150 is thought[5] to have a structural role, connecting HSP101 and EXP2.

Reduced expression[7] of HSP101 or PTEX150, or inhibition[8] of the assembly of HSP101 into the PTEX complex, results in parasite death. PTEX is specific to species of the genus *Plasmodium* and is not made by humans. It is an attractive drug target because it provides the only known route by which parasite proteins enter the cytoplasm of a red blood cell. However, PTEX's relative

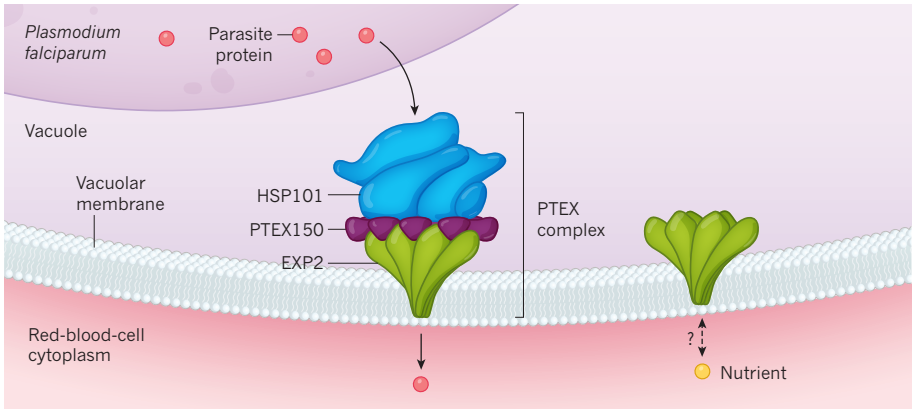


Figure 1 | Proteins that are essential for the malaria parasite's survival. During part of its life cycle, the *Plasmodium falciparum* parasite resides inside a membrane-bound vacuole in human red blood cells. Ho *et al.*[2] report the structure of a parasitic multiprotein complex called PTEX, which is present on the inside of the vacuolar membrane. This complex is essential for parasite survival, and proteins are exported from the vacuole through PTEX into the cytoplasm of red blood cells. The authors' analysis provides a detailed view of three of the proteins that form PTEX: EXP2, HSP101 and PTEX150 (only 20% of the structure of PTEX150 was determined). The authors' work also illuminates how proteins transit through the PTEX complex. Garten *et al.*[4] report that EXP2 can form a channel that enables nutrients to be transported across the vacuolar membrane. Whether this occurs in both directions, or in only one, is not known.

novelty offers few clues to how it functions. EXP2 synthesized in the laboratory can form protein channels in lipid bilayers[9]. However, there have been no reports of full-length HSP101 or PTEX150 having been successfully synthesized for use in *in vitro* experiments. This has prevented structural analysis of the proteins, or reconstitution of the core PTEX complex in lipid membranes, to determine how the complex assembles and functions.

Because of these experimental limitations, Ho *et al.*[2] opted instead to extract PTEX directly from red blood cells containing the parasite. Then, using a technique called cryo-electron microscopy (cryo-EM), the authors captured two distinct structural conformations of the core PTEX complex in the process of exporting unfolded protein cargo — they called these conformations the 'engaged' and 'resetting' states. The cryo-EM analysis revealed that HSP101, PTEX150 and EXP2 assemble into an asymmetrical structure containing six molecules of HSP101, seven of PTEX150 and seven of EXP2. These structures closely align with models of the organization and size of PTEX that had been predicted from biochemical and protein-analysis experiments[3,5].

Ho and colleagues found that the seven EXP2 molecules, which make up the protein channel in the lipid membrane, create a funnel shape, with the amino terminus of each molecule forming a transmembrane helix in the vacuolar membrane to provide an anchoring 'stem' (Fig. 1). The 'mouth' of EXP2 constitutes the bulk of the protein, and faces into the vacuole. This end of EXP2 contains a domain that tethers it to the carboxy-terminal domain of HSP101, situated directly on top. Only approximately 20% of the structure of PTEX150 could be determined. Nevertheless, this was sufficient to reveal that each PTEX150 molecule slots in between adjacent EXP2 molecules at the mouth of the EXP2 funnel, curling down towards the stem. Thus, PTEX150 provides a protective path for unfolded protein cargo transiting from HSP101 to EXP2.

Of the three proteins, HSP101 displayed the greatest structural difference between the engaged and resetting states of PTEX, and on this basis the authors propose a mechanism for how cargo is threaded through PTEX's central cavity. In this model, domains of the six assembled HSP101 molecules form two 'hands' that work together to thread unfolded cargo through the PTEX150 and EXP2 funnel. In the engaged state of PTEX, both the 'active' and 'passive' hands of HSP101 grasp the unfolded cargo. The cargo is then fed downwards through the central cavity of PTEX in a spiral fashion as it passes from the active to the passive hand. In the resetting state, HSP101's active hand moves upwards to grasp the next section of the cargo protein for transport, and the passive hand grips the cargo to prevent it from slipping backwards and away from the PTEX channel.

The cryo-EM structures provide insight into several crucial interactions between the PTEX components. These interactions are potentially required for assembly and optimal function of the complex, and could be tested using genetic approaches to validate the model. Ho and colleagues were unable to determine the structure of the N-terminal domain of HSP101 that binds the protein cargo. Thus, it is unclear how cargo is recognized by HSP101, and whether cargo proteins are unfolded by proteins known as chaperones before they reach PTEX. Given that unfolded proteins pass through PTEX, these cargo proteins would then need to be refolded to function, presumably by other chaperone proteins. However, because EXP2 does not

extend into the cytoplasm of red blood cells, it is unclear how chaperone proteins in the host cell might be recruited to cargo exiting PTEX.

Garten *et al.*[4] investigated EXP2 using *in vitro* experiments, and report that it has another role in addition to its function in PTEX. Previous experiments using electrophysiological techniques have shown that a channel exists in the vacuolar membrane of parasite-infected red blood cells through which nutrients such as amino acids and sugars can pass[10], but the identity of this channel has been a mystery. In electrophysiological studies, Garten and colleagues demonstrated a direct relationship between the level of expression of EXP2 and the frequency of detection of the mysterious channel. When the authors generated a version of EXP2 that had a truncated C-terminal domain, which is located in the vacuole and is not required for protein export, this altered the voltage-response properties of the nutrient channel, leading the authors to conclude that EXP2 is indeed the elusive nutrient channel.

That EXP2 might have a role separate from its function in PTEX is consistent with evidence that EXP2's gene-expression profile differs from that of the other PTEX components[5]. Moreover, the authors found that most EXP2 is not present in a complex with PTEX. Although EXP2 is essential for parasite survival[11,12], the contribution of the EXP2 nutrient channel to parasite growth remains unknown. The channel could be characterized in detail if EXP2 was incorporated into lipid bilayers for *in vitro* experiments.

The studies by Ho, Garten and their respective colleagues offer a close look at how major *P. falciparum* proteins function. Interestingly, EXP2 is evolutionarily conserved among vacuolar-dwelling parasites called apicomplexans[1]. Perhaps the nutrient-transiting capacity of EXP2 was adapted by *P. falciparum* to generate a protein-conducting channel that evolved through the recruitment of other proteins such as HSP101 and PTEX150. EXP2 and PTEX are expressed throughout the life cycle of *P. falciparum*, so drugs that target them might be highly effective at tackling malaria. These new insights into the interactions between the components of PTEX offer exciting possibilities for the development of peptides or small molecules that might block the function of this complex. ■

Tania F. de Koning-Ward *is in the School of Medicine, Deakin University, Geelong, Victoria 3216, Australia.*
e-mail: tania.dekoning-ward@deakin.edu.au

1. de Koning-Ward, T. F. *et al. Nature Rev. Microbiol.* **14**, 494–507 (2016).
2. Ho, C.-M. *et al. Nature* **561**, 70–75 (2018).
3. de Koning-Ward, T. F. *et al. Nature* **459**, 945–949 (2009).
4. Garten, M. *et al. Nature Microbiol.* http://doi.org/10.1038/s41564-018-0222-7 (2018).
5. Bullen, H. E. *et al. J. Biol. Chem.* **287**, 7871–7884 (2012).
6. Johnson, D. *et al. Parasitology* **109**, 1–9 (1994).
7. Elsworth, B. *et al. Nature* **511**, 587–591 (2014).
8. Beck, J. R., Muralidharan, V., Oksman, A. & Goldberg, D. E. *Nature* **511**, 592–595 (2014).
9. Hakamada, K., Watanabe, H., Kawano, R., Noguchi, K. & Yohda, M. *Biochem. Biophys. Res. Commun.* **482**, 700–705 (2017).
10. Desai, S. A., Krogstad, D. J. & McCleskey, E. W. *Nature* **362**, 643–646 (1993).
11. Matthews, K. *et al. Mol. Microbiol.* **89**, 1167–1186 (2013).
12. Matz, J. M., Matuschewski, K. & Kooij, T. W. A. *Mol. Biochem. Parasitol.* **191**, 44–52 (2013).

This article was published online on 27 August 2018.

QUANTUM PHYSICS

Designer atom arrays for quantum computing

A key step in the development of quantum computers that use neutral atoms as quantum bits is the assembly of tailored 3D arrays of atoms. Two laser-based approaches have now been reported to do this. SEE LETTERS P.79 & P.83

NATHAN LUNDBLAD

Quantum computers and simulators are of enormous interest because of their potential to shed light on mysteries of physics that are difficult to model using conventional computers. Some physical platforms used in realizing quantum-computing protocols — including trapped ions and several solid-state systems based on superconductors — have received increased attention in the past year. But in this issue, two groups report technical breakthroughs that will aid the development of another platform: trapped neutral atoms. Barredo *et al.*[1] (page 79) report their use of precision optical-engineering methods to sort atoms into arbitrary 3D patterns, whereas Kumar *et al.*[2] (page 83) construct cubic lattices by revisiting a fanciful thought experiment known as Maxwell's demon. The ability to organize neutral atoms exactly into planned 3D arrays will be valuable for the development of neutral-atom quantum computers that use a large number of quantum bits (qubits).

Arrays of isolated neutral atoms have long shown promise for quantum computing because neutral-atom qubits are extremely well isolated from environmental noise and are highly controllable, and also because such systems can be scaled up to large numbers of qubits[3,4]. Given that controlled interactions between atoms are needed to perform quantum-computing operations, neutral-atom quantum computers will need qubits to be precisely arranged in a specified pattern. However,

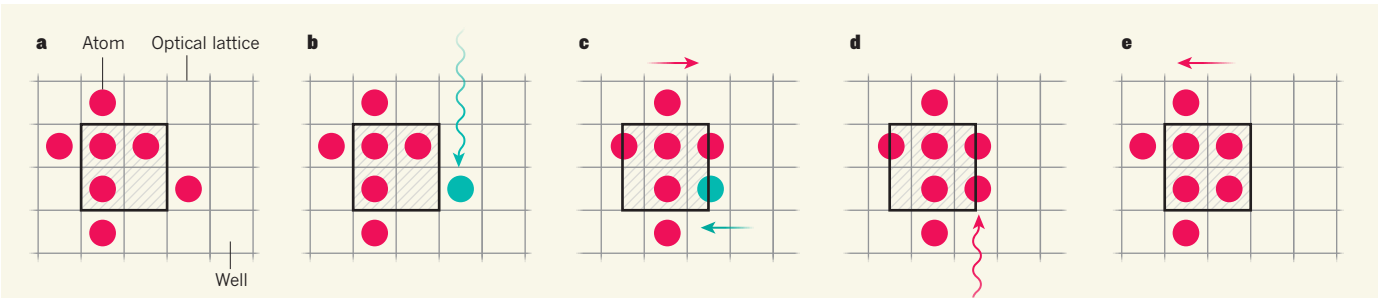


Figure 1 | A protocol for arranging neutral atoms in cubic optical lattices. Kumar *et al.*[2] report a method for arranging ultracold, neutral caesium atoms in defined patterns in a cubic, 3D optical lattice — a series of laser-generated potential-energy wells in which atoms can be confined. Only one layer of atoms is shown, for simplicity. **a**, The atoms start off in random positions and in the same electronic state (state A, red). The shaded square indicates a target region that is to be filled with atoms. **b**, A combination of lasers and microwaves (wavy arrow) flips the state of one atom into a different state (state B, turquoise). **c**, A lattice shift is induced that moves the lattice and all atoms in state A half a step to the right and those in state B half a step to the left. **d**, The atom in state B is flipped back to state A. **e**, A reverse lattice shift moves the lattice and all atoms in state A half a step to the left, so that the square region is now filled with atoms.

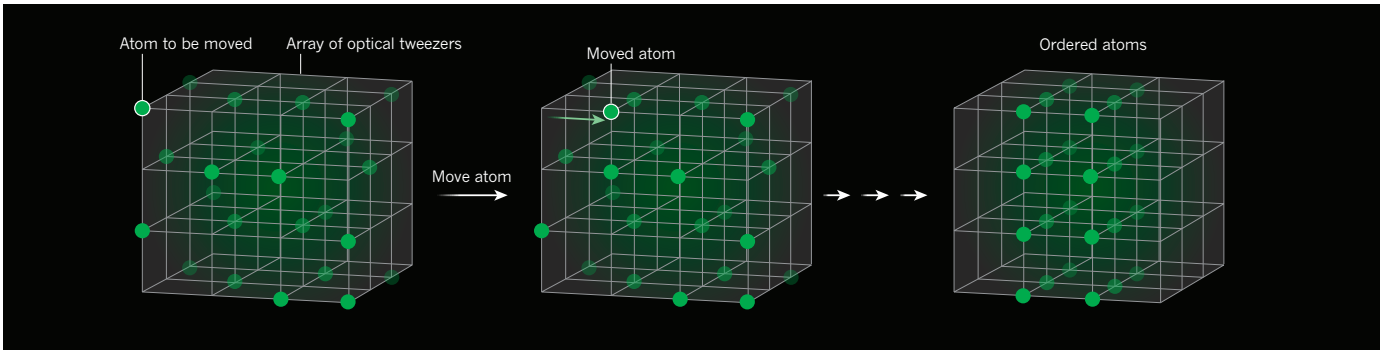


Figure 2 | A protocol for arranging neutral atoms in arbitrary 3D patterns. Optical tweezers are laser-generated optical traps that can capture atoms. Barredo *et al.*[1] formed 3D arrays of optical tweezers in arbitrary patterns (each vertex of the array represents an optical tweezer), and part-filled them with ultracold rubidium atoms, which initially reside at random positions. The authors then used a movable optical tweezer (not shown) to grab atoms at 'incorrect' positions and deposit them at desired positions, to produce precise arrangements of atoms.

developing methods for sorting atoms into patterns has proved challenging. Neutral-atom qubits require ultracold temperatures and extremely high vacuums to function, and therefore require complicated apparatus; ordering them into arrays using optical techniques adds an extra level of practical complexity. Progress has been made in arranging neutral atoms in one and two dimensions[5–8], but 3D stacking will become essential as the number of qubits used approaches the hundreds, or to construct arrangements that have topologies not achievable in two dimensions.

Kumar *et al.* have extended their previously reported approach[9] to assemble cold clouds of caesium atoms into a 3D lattice. The method begins with a randomly populated optical lattice: a trap formed from the interference patterns of counter-propagating lasers, in which atoms can be confined much like eggs in cartons. After imaging and recording the random locations of atoms in the lattice, the authors implement a sorting protocol that involves intricately controlling the polarizations of the lattice lasers, while using additional 'addressing' lasers and microwaves to position any given atom within a $5 \times 5 \times 5$ array of lattice sites (Fig. 1). In this way, up to 50 neutral atoms can be precisely ordered into an array that is suitable for use in a quantum computer.

Kumar *et al.* frame their sorting and preparation protocol in terms of Maxwell's demon. This thought experiment was proposed by James Clerk Maxwell in 1867, and explores the nature of entropy, a measure of disorder. Maxwell postulated that a reversible sorting mechanism (a sentient demon, although a non-sentient process would also work) could partition gas molecules into two sub-volumes. But this sorting process would lower the entropy of the gas in apparent violation of the second law of thermodynamics, which states that the entropy of any isolated system can only increase. How can this conundrum be explained? The answer turns out to be that the act of sorting inevitably increases the entropy of the Universe. Because the dominant entropy in Kumar and colleagues' experiments is associated with the physical arrangement of the atoms, their work is a realization of an omniscient Maxwell's demon, summoned to organize the initial arrangement of a qubit array.

Meanwhile, Barredo *et al.* extend their previously reported method[10] for 2D atom sorting to three dimensions. Their approach to disorder and sorting is different from Kumar and colleagues' method, but just as effective. They use a holographic technique whereby a laser beam is reflected off a spatial light modulator and then focused to form traps known as optical tweezers. In this way, they generate arrays of traps in arbitrary configurations that can be loaded with up to 72 cold rubidium atoms. To remove disorder and build the desired atomic configuration, the authors use a separate, movable optical tweezer to pluck atoms from 'wrong' traps and either move them to correct sites or discard them (Fig. 2). This allows them to build qubit arrays in standard grid patterns, in topologies such as a Möbius strip, and even in the shape of the Eiffel Tower (see Fig. 2 of the paper[1]).

Barredo and colleagues go on to engineer an interaction between two qubits in a sorted array. To do this, they excite the atoms into 'Rydberg' states, which produce atomic electrical dipoles that allow the qubits to sense each other through dipole–dipole interactions. By contrast, atoms in their ground states have vanishingly small dipole–dipole interactions. Rydberg interactions have previously been used to enable quantum-logic operations carried out by small systems of neutral-atom qubits[3,4], and could form the basis of both the current groups' future efforts to develop quantum computers.

The two papers report similar milestones for the assembly of neutral-atom quantum computers, with Barredo and colleagues also reporting a working two-qubit interaction. However, the atoms in Barredo and colleagues' system are not as cold as they could be, which means that the entropy remaining in their arrays is substantially greater than in Kumar and colleagues' system. The resulting micrometre-scale motion of the atoms within the traps could limit the performance of future devices based on this system — a restriction that does not apply to Kumar and co-workers' apparatus. But Barredo and colleagues' approach does allow qubit arrays of any spatial design to be made, whereas Kumar and co-workers' apparatus generates only a cubic lattice. These differences might not be important for near-term quantum-computing goals, however. It remains to be seen whether quantum entanglement (a phenomenon that produces stronger correlations between particles than those permitted by classical physics, and which fuels quantum-computing algorithms) can be created for such large numbers of working qubits.

Both papers report technical tours de force, and showcase how far neutral-atom systems have come in terms of stability, reproducibility and technical sophistication. The next step is probably to generate quantum entanglement between arbitrary pairs of atoms in sorted arrays. It will also be interesting to see which exotic quantum states can be simulated using these qubit arrays, especially if some of those states cannot be modelled using existing computational approaches[11]. Finally, it will be exciting to see whether the potential advantages of neutral atoms will now begin to pay dividends in the race to develop a working quantum computer. ■

Nathan Lundblad *is in the Department of Physics and Astronomy, Bates College, Lewiston, Maine 04240, USA.*
e-mail: nlundbla@bates.edu

1. Barredo, D., Lienhard, V., de Léséleuc, S., Lahaye, T. & Browaeys, A. *Nature* **561**, 79–82 (2018).
2. Kumar, A., Wu, T.-Y., Giraldo Mejia, F. & Weiss, D. S. *Nature* **561**, 83–87 (2018).
3. Weiss, D. S. & Saffman, M. *Phys. Today* **70**, 44–50 (2017).
4. Saffman, M. *J. Phys. B* **49**, 202001 (2016).
5. Endres, M. *et al. Science* **354**, 1024–1027 (2016).
6. Robens, C. *et al. Phys. Rev. Lett.* **118**, 065302 (2017).
7. Lester, B. J., Luick, N., Kaufman, A. M., Reynolds, C. M. & Regal, C. A. *Phys. Rev. Lett.* **115**, 073003 (2015).
8. Sturm, M. R., Schlosser, M., Walser, R. & Birkl, G. *Phys. Rev. A* **95**, 063625 (2017).
9. Wang, Y., Kumar, A., Wu, T.-Y. & Weiss, D. S. *Science* **352**, 1562–1565 (2016).
10. Barredo, D., de Léséleuc, S., Lienhard, V., Lahaye, T. & Browaeys, A. *Science* **354**, 1021–1023 (2016).
11. Gross, C. & Bloch, I. *Science* **357**, 995–1001 (2017).

REVIEW

https://doi.org/10.1038/s41586-018-0457-8

Facing up to the global challenges of ageing

Linda Partridge[1,2]*, Joris Deelen[1,3] & P. Eline Slagboom[1,3]*

Longer human lives have led to a global burden of late-life disease. However, some older people experience little ill health, a trait that should be extended to the general population. Interventions into lifestyle, including increased exercise and reduction in food intake and obesity, can help to maintain healthspan. Altered gut microbiota, removal of senescent cells, blood factors obtained from young individuals and drugs can all improve late-life health in animals. Application to humans will require better biomarkers of disease risk and responses to interventions, closer alignment of work in animals and humans, and increased use of electronic health records, biobank resources and cohort studies.

During the last 200 years, average human life expectancy has doubled in most developed countries[1] (Fig. 1). Better quality of water, food, hygiene, housing and lifestyle, immunization against infectious disease, antibiotics and improved medical care first reduced mortality in early life[2,3] and, after about 1950, in people of 70 years of age or older[1,4]. Whether there will be a limit to human life expectancy is vigorously debated[5], but survival rates in the elderly and mean life expectancy are generally projected to continue to increase[6]. In parallel with longer lives, most aspects of age-specific health have also improved, with increases in both physical and cognitive functioning during ageing in successive birth cohorts[7,8].

Recent increases in human life expectancy have been much too rapid for genetic change to have had more than a minor role[9]. In contemporary populations, individuals who survive to great ages are particularly common in the so-called 'blue zones' of the world, Okinawa in Japan, part of Sardinia in Italy, Ikaria in Greece, Nicoya in Costa Rica and Loma Linda in the United States. These populations have not been found to be genetically distinct from their neighbours, and environment and lifestyle, including social networks, seem to have important roles in the healthy ageing of these people[10]. Factors such as diet, education and physical activity throughout postnatal life have a cumulative effect on mortality[11], and conditions during early life and parental health also have a large influence[3].

Improved health of people of all ages, including older people, and the consequent increase in life expectancy, are to be celebrated as achievements of civilization. However, healthy, disease-free lifespan (healthspan) has not increased as much as lifespan[12]. A global increase of five years in total life expectancy between 2000 and 2015 has been accompanied by only 4.6 years of healthy life expectancy (see http://apps.who.int/gho/data/view.main.SDG2016LEXREGv?lang=en). An average 16–20% of life is now spent in late-life morbidity[13], longer in females than in males, and in individuals with a lower socio-economic status or obesity[13–15]. Most of us now live far longer than in our evolutionary past, to ages that have not been shaped by natural selection. Advancing adult age is therefore the major risk factor for chronic killer diseases, including cancer and cardiovascular and neurodegenerative diseases[16] (Fig. 2). The burden of these conditions is now falling mainly on older people. Ageing impairs sensory, motor and cognitive function, and thus lowers quality of life. Reduction in the length and severity of late-life morbidity should therefore be a major aim in civilized societies in the future. We shall refer to this goal as 'compression of morbidity'.

Compression of morbidity should be achievable. First, individuals who survive to over 100, 105 or 110 years show progressively greater compression of late-life morbidity[17,18]. Therefore, a relatively healthy end to life is physiologically feasible and, if we could find the underlying mechanisms, it might be possible to extend the trait to the general population. Second, experimental work with laboratory animals, mainly yeast, nematode worms, fruitflies and mice, has revealed the remarkable malleability of ageing. Genetic, environmental and pharmacological interventions can extend lifespan, ameliorate the loss of function and diseases of ageing and, in some cases, compress late-life morbidity[19–21]. Although laboratory animals do not live as long as humans, ageing has underlying mechanisms that are conserved over long evolutionary distances, and these provide potential targets to maintain human health at older ages[22]. Indeed, similar life-extending interventions are effective in different laboratory species[19,21].

Here, we address the opportunities and challenges for discovering the genetic and environmental determinants of human lifespan and healthspan, and in translating results of discoveries in animals into health improvements for ageing humans. We will not to be able to abolish ageing, but we do expect to be able to attenuate the process and greatly ameliorate its effects.

Genetics of human lifespan and healthspan

Genetic analysis of the marked individual variation in human lifespan could identify potential targets for intervention, and several approaches have been used (Box 1 and Table 1). Twin studies have suggested that human lifespan is around 25% heritable[23], indicating that there is a large, and possibly modifiable, effect of environmental factors on lifespan. A recent study[24] in a population of millions of individuals, using the population pedigree, showed an even lower heritability of only 12%. The variation in these figures is probably due to the difficulty of accurately estimating common environmental and behavioural effects within families. The heritability of lifespan is minimal for parents who die between puberty and the age of 60, and then increases progressively with death at later ages[25]. Different measures, including overall lifespan, healthspan and survival to exceptionally old ages (often termed longevity), have been used in genetic studies. Multiple genome-wide association studies (GWAS) of longevity have been performed, and the only genetic locus to show robust, genome-wide significance across studies is apolipoprotein E (*APOE*)[26], a cholesterol carrier in

[1]Max Planck Institute for Biology of Ageing, Cologne, Germany. [2]Institute of Healthy Ageing, and Department of Genetics, Evolution and Environment, UCL, London, UK. [3]Molecular Epidemiology, Department of Biomedical Data Sciences, Leiden University Medical Center, Leiden, The Netherlands. *e-mail: linda.partridge@age.mpg.de; p.slagboom@lumc.nl

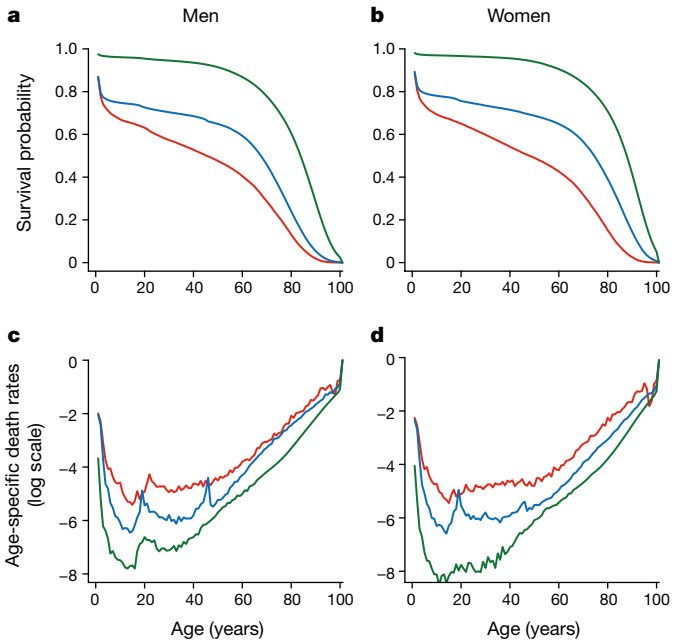


Fig. 1 | Cumulative survival and age-specific death rates in the Netherlands in 1850, 1900 and 1950. a–d, Cumulative survival (**a**, **b**) and mortality rates (**c**, **d**) in men (**a**, **c**) and women (**b**, **d**) based on 100,000 individuals per birth cohort (1850 (red), 1900 (blue) and 1950 (green)) from life tables from the Netherlands. **c**, **d**, Note that the *y* axis is a log scale.

peripheral tissues and the brain that is also associated with the susceptibility to cardiovascular and Alzheimer's disease[27] (Box 1). The general lack of replication of findings in independent studies may be attributable to different measures of survival and health, the age specificity of genetic effects[28], and different allele frequencies of lifespan-associated genetic variants in different birth cohorts[29].

Survival to advanced ages, particularly the 1–10% longest lived of the generation, is enriched in families[23], and members of these families show a lifelong survival advantage, with lower risk of coronary artery disease, cancer and type 2 diabetes[30–32] and better immune and metabolic health in middle and old age[32–34] compared to the general population and even to their spouses. However, the effects of common, non-genetic influences in early life in these families cannot be ruled out. Neither familial nor sporadic long-lived individuals[35,36] display a decreased load of common genetic risk variants for age-related disease[36,37]. However, any common protective genetic variant that is responsible for familial longevity has yet to be found. Replicated studies based on candidate genes emerging from studies in model organisms can also be informative[38,39], and these have identified the *FOXO3A*[40] locus, which encodes a transcription factor, the homologues of which, in model organisms, have a consistent role in healthy ageing[19,41,42]. Future genetic studies of longevity could focus both on establishing the longevity phenotype in multi-generational families, with whole-genome or exome sequencing of targeted cases, and on the multimorbidity phenotype itself, rather than on proxies such as healthspan.

Phenotypes and mechanisms of human ageing

Human mortality rates reach a minimum around puberty and increase roughly exponentially thereafter (Fig. 1). Initially, the ageing process is manifested sub-clinically, in various types of physiological deterioration. From the third decade onwards, age-related changes in body composition occur, including loss of bone, cartilage, muscle mass and strength and gain of abdominal fat[43,44]. Subsequently, systemic changes occur, for instance, in the endocrine system, resulting in altered hormone levels, and in the circulation, resulting in changes in blood pressure and blood lipids. The responses of tissues to hormones can also be affected, as in insulin resistance[45]. Mechanical and structural changes also occur, including vascular stiffness, which can affect heart and brain functions[46]. Eventually, these continuous sub-clinical changes can culminate in a range of medically defined disease conditions in middle age, with the co-existence of two or more chronic health conditions in an individual being defined as multimorbidity. People with higher levels of markers of disease risk in their blood (Box 2), and those with multimorbidity, die up to 20 years younger than those with lower levels[47,48]. Late ages are frequently accompanied by frailty[49], a composite index of ill health, functional and psychosocial deficits[50], which increases the risk of falls, fractures, hospitalization, organ failure, disability and death[51] (Fig. 3).

The largest medical challenges in treating the growing number of elderly patients are multimorbidity, present in at least half of the elderly over 70 years[47,52] and the related use of five or more types of medication (polypharmacy), which occurs in over 10% of the general population[53,54] and 30% of the elderly[55]. Up to 12% of all hospital admissions of older patients can be attributed to adverse drug reactions[56,57], which most commonly involve anticoagulants, blood pressure lowering and hypoglycaemic drugs, antiplatelet agents (aspirin) and nonsteroidal anti-inflammatory drugs[58]; the latter two contribute most frequently to death after admission. Behavioural risk factors also have a major role. Large multi-cohort studies in high-income countries have indicated that the number of years lost because of smoking, physical inactivity and high alcohol intake (more than 21 units per week for men, more than 14 per week for women) are, on average, 4.8, 2.4 and 0.5 years, respectively[59]. Sedentary behaviour is especially common among older people, who spend, on average, almost 10 waking hours in an immobile posture[60]. The WHO (World Health Organization) is therefore targeting the major risk factors that have been identified so far, with the overall aim of reducing premature mortality from non-communicable diseases by 25% by 2025[61].

Further progress in preventing late-life ill health will come from better predictors of its occurrence, and from understanding how to intervene to block the causal mechanisms at an early stage. Different measures (biomarkers) can indicate the aetiology of ageing and its progress to disease states (Box 2). Physiological decline can partly be measured by standardized analyses of physical, respiratory and cognitive capacity, blood pressure and circulatory markers. Poor scores for these indicators during midlife are associated with an increased risk of morbidity and mortality over time[11]. These markers can also monitor health improvement in response to interventions, but do not yet robustly reflect all relevant aspects of ageing. Generating comprehensive biomarker profiles that are capable of doing so is therefore important, and this field is progressing rapidly (Box 2).

Preventative interventions into lifestyle aimed at slowing specific effects of ageing have presented a complex picture, with outcomes varying with the type of intervention, the age of the subjects and the population from which they are drawn. Some intervention regimes have been successful. For instance, treating adults at risk of diabetes by altering their diet, increasing their physical activity or both can be as effective as medication, and with better continued benefit, even for up to 15 years[62]. Reductions in hypertension, diabetes and brain atrophy, improved cognitive performance and reduction in mortality due to cancer and cardiovascular disease, have all been achieved by alterations to lifestyle. Specific diets[63,64], exercise[65], the two combined[66], cognitive training and vascular risk management[67], caloric restriction[68], intermittent fasting[69] and supplementation of vitamin D[70] have all been reported to be effective for specific conditions. However, the response to these interventions can show marked individual variation. It may become possible to target these interventions to those individuals who will benefit the most when robust biomarkers of the variation become available. For instance, older and more frail people could benefit from more dietary protein to combat traits such as muscle wasting and weakness (sarcopenia)[71], whereas middle-aged people may benefit from less protein to combat cancer, although more direct evidence is needed, both from experimental work in animals and epidemiological studies in humans.

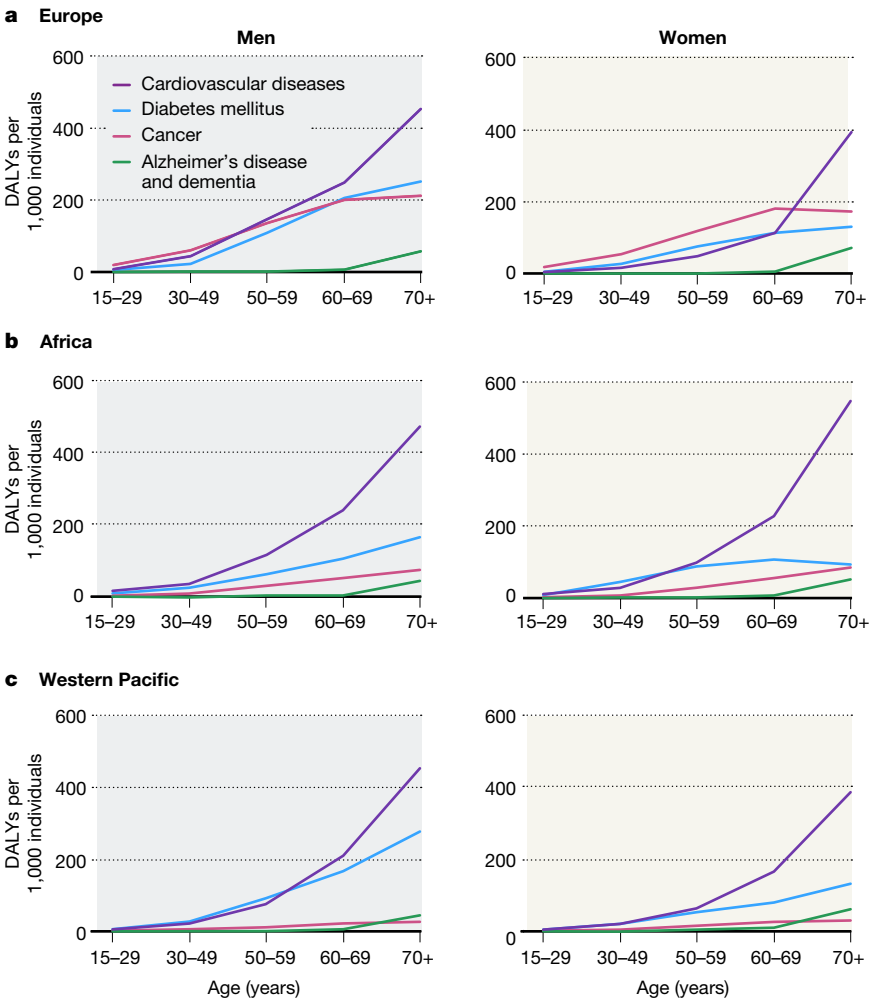


Fig. 2 | Disability-adjusted life years for age-related diseases in three global regions. a–c, Disability-adjusted life years (DALYs) per 1,000 individuals are shown for men (left) and women (right). DALYs are shown for malignant neoplasms (red), diabetes mellitus (blue), Alzheimer's disease and other dementias (green) and cardiovascular diseases (purple) in different areas of the world for the year 2015 according to data from the WHO[165]. **a**, Europe. **b**, Africa. **c**, Western Pacific. One DALY represents the loss of one year of health due to mortality or disability caused by the indicated disease.

Lifestyle interventions, while often beneficial, can be insufficient to prevent the progress of age-related problems, partly because of failures in compliance, and also because of limited and variable responses. Drugs are an additional option, and are already in widespread use for the prevention of cardiovascular disease, by pharmacologically decreasing hypertension[72] and low-density lipoprotein cholesterol[73,74] in healthy individuals who are at risk of cardiovascular disease (primary prevention). Treatment of the elderly is complex, since the relation between cardiovascular risk indicators, such as high body mass index, blood pressure and blood lipids, and end points, such as mortality, can change and even reverse with increasing age. The changing correlation with age could indicate that pharmacological interventions should depend on age[75] and the presence of frailty and multimorbidity[76]. However, mortality may be selective, with those sensitive to classical risk factors dying before the age of 70, or reverse causation may occur, with age-related diseases leading to low body mass index and blood pressure, and further work teasing out causality is needed. The ageing process in animals shows evolutionarily conserved, parallel and interacting mechanisms, known as hallmarks[22], that have proven to be modifiable, and several of these are also well-documented in humans (Table 2). They eventually lead to unrepaired damage in DNA[77], accumulation of misfolded and aggregated proteins (for example, in the brain and the retina) and senescent cells (for example, in joints and kidneys)[78] as well as to an inappropriate and persistent activation of stress responses[79], such as in the innate immune system (inflammaging[80]).

To develop further interventions to compress morbidity, including drugs, we need a better understanding of the roles of individual ageing mechanisms in different tissues and at different stages in life, and their contributions to the aetiology of age-related diseases. To this end, animal studies are useful to inform more targeted studies in humans.

Translating discoveries from animal ageing

The model organisms that are commonly used in ageing research have a much shorter lifespan than humans. However, they recapitulate many features of human ageing[81]. Furthermore, similar to many humans, their culture and care regimes in the laboratory mostly protect them from infectious diseases, provide them with abundant high-quality food, restrict their exercise and remove many physical challenges. As a consequence they, too, live to much greater ages than in their evolutionary past. Conservation of mechanisms of ageing between animals and humans extends to both the hallmarks[22] of ageing and the genes that are involved in ageing and age-related diseases[16,20]. Different model organisms best recapitulate specific aspects of human age-related problems. Work across different model organisms has yielded biomarkers (Box 2) that predict remaining lifespan, such as nucleolar volume[82] and telomere length[83,84], and these are promising candidates for inclusion in a multivariate predictor of the rate of human ageing.

Ageing in animals has proved to be highly malleable in response to environmental and genetic interventions. Various regimes of dietary restriction are particularly effective, with increased lifespan and

Box 1

Genetic studies into the variation in human lifespan

Studies aiming to identify genes that influence human lifespan initially explored candidate genes related to human age-related diseases or to amelioration of ageing in laboratory animals. After gene array-based technologies became available, genetic variation in the whole genome was explored in linkage studies of longevity families or GWAS of population-based older cases and younger controls. The selection of cases for these studies was based on longevity threshold criteria, such as survival of individuals to ages above 90 or 100 years or individuals (or their parents) belonging to the top 10% or 1% of survivors in a population[141,142,145–147], or a continuous lifespan parameter, such as age at death of individuals or their parents[143,144,166]. The candidate gene studies revealed only two loci that have been consistently replicated in independent studies. The first is *APOE*, which is also the only genetic locus that shows robust, genome-wide significance across different GWAS (Table 1). The genetic variant responsible for ApoE ε2 (rs7412) has been shown to be protective and the one for ApoE ε4 (rs429358) has been shown to be deleterious[167]. The second is *FOXO3A*[41,42] (see LongevityMap[168] for an overview of results from all candidate gene studies). Genomic locations that were identified in linkage studies showed no overlap between the different studies[26]. The GWAS analyses, on the other hand, have thus far identified several genetic variants (Table 1). However, the majority of these are disease-related variants that influence early mortality, rather than survival to extremely old ages. One of the loci that was observed only in a single, large, parental longevity study was also found to be associated with a diversity of age-related diseases[169], and contained *CDKN2A* and *CDKN2B*, which are involved in the development of cellular senescence, a hallmark of ageing. Individuals from long-lived families may have genetic variants that are rarer, which can only be identified by (whole-genome) sequencing of family members. This approach has already resulted in the identification of functional genetic variants in the IGF-1 receptor (*IGF1R*)[38,170].

a remarkably broad improvement in health during ageing in diverse species, including rodents[20,68,85]. Two studies[86,87] of dietary restriction in rhesus monkeys had slightly different outcomes, probably because of differences in the composition of the control diet, the degree of restriction and the timing of food provision[88]. Lifespan increased with dietary restriction in one study, while in both there were major improvements in health in food-restricted animals, with reduced plasma triglycerides, diabetes, cardiovascular disease, sarcopenia, incidence of neoplasms and brain atrophy, which are the most relevant health parameters in ageing humans. Multiple genetic interventions can also induce broad improvements in health in laboratory animals[89]. For example, reduced activity of the insulin–insulin-like growth factor (IGF) signalling (IIS)–mammalian target of rapamycin (mTOR) signalling network can extend lifespan in yeast, worms, fruitflies and mice[21], and genetic variants in candidate orthologous genes, or their gene expression patterns, in humans can be associated with survival to advanced ages[41,42,90–92]. As for dietary protein, any benefits of modulating the activity of the IIS–mTOR signalling network in humans may depend on age[71]. The network detects nutrition and stresses, and matches costly activities such as growth, metabolism and reproduction to current physiological state. Systems mediating major life history choices in response to environmental cues thus have an important role in ageing. Despite the complexity of the ageing process, with multiple hallmarks and interactions between them, its effects can clearly be ameliorated in animals. Notably, these interventions have also proven to be capable of combatting the pathology in models of age-related diseases, including cancer, neurodegeneration and cardiovascular diseases[93–95]. It remains to be seen whether human ageing will show a similar degree of adaptability in response to interventions that have proven to be effective in animals, but a major conclusion from ageing in model organisms is that delaying, or even preventing, age-related diseases is a realistic prospect. Furthermore, while no intervention studied so far has improved all aspects of health[96], interventions that extend lifespan generally also prevent more than one age-related condition simultaneously[95] (Fig. 4).

Anti-ageing interventions that have proven to be effective in laboratory animals are starting to be assessed for their feasibility, effectiveness and safety in humans. Although dietary restriction increases predictors of healthy ageing in human volunteers[68], it is not a realistic public health intervention, because of poor compliance with even mild (90%) dietary restriction regimes[97]. However, more modest dietary interventions may be realistic. For example, in mice modulation of the protein content of the diet can both prevent over-consumption of protein-poor diets and avoid the increased risk of cancer from high-protein diets[71,98]. The amino acid content of protein also determines its value to the animal and can be modulated to reduce total protein consumption[99]. Timing of food intake can also be important[100,101]. Food-restricted mice and rats are usually fed once a day and consume their limited food ration as soon as it is supplied, with a protracted fasting period until the next day. This fasting period may be at least as important as reduced food intake in promoting healthy ageing[98,102,103]. Indeed, trials in middle-aged humans with a fasting-mimicking diet, low in protein, carbohydrate and calories, but high in unsaturated fats, have shown beneficial effects on biomarkers of health, such as blood pressure and levels of circulating IGF-1, with particularly strong effects in individuals who are most at risk of disease[104]. The time of day at which food is consumed can also have substantial metabolic effects[105]. These more nuanced interventions need further exploration in both animal and human studies, particularly with regards to the age specificity of their effects and possible adverse effects at later ages.

Increasing attention is focusing on pharmacological manipulation of the mechanisms of ageing in animals, with a view to direct translation to humans to prevent age-related diseases. Development of new drugs to ameliorate human ageing would pose challenges for clinical trials since, in the absence of validated biomarkers of risk, a large, random, initially healthy population would have to be treated over a long period. At present, repurposing of existing drugs with a good safety profile is therefore a more realistic short-term prospect than de novo drug development[106]. Because the mechanisms of ageing that have been discovered in animals are proving to be important for human age-related diseases, many are already targets of drugs that are licensed to treat these diseases[107,108]. There is an opportunity to widen the use of existing drugs that are used to treat single, age-related diseases to prevent multimorbidity. For instance, the licensed drug sirolimus (also known as rapamycin) inhibits mTOR complex 1, part of the nutrient- and stress-sensing network, and can extend the lifespan of model organisms, including mice, in which it improves many, but not all, aspects of health during ageing and protects against cancer[93,109]. As in elderly mice, the poor immune response of elderly humans to immunization against influenza can be enhanced by pretreatment with the related drug everolimus[110]. The anti-diabetic drugs metformin and acarbose can also extend lifespan in laboratory animals, and are currently registered for clinical trials against ageing itself, which has not previously been recognized as a valid target[106,111–113]. The doses of drugs that are effective for the prevention of the effects of ageing in animals are often lower than those used clinically, so that side effects may be reduced, and may be further prevented by making drugs, such as rapamycin, more specific to their therapeutic target[114] and by adjusting dosing regimes[114,115].

Other recent discoveries about animal ageing are showing promise for translation to humans. Cellular senescence, a hallmark of ageing in both laboratory vertebrates and humans (Table 2), is a permanent type of cell cycle arrest and is associated with resistance to cell death

Table 1 | Loci emerging from GWAS of discrete and continuous lifespan-related phenotypes in human studies

			Replication		
Closest gene(s)	**Discrete phenotypes**	**Continuous phenotypes**	**Within publication**	**Between publications**	**Associations with age-related diseases**
APOE[141–145]	Age ≥99th percentile;age ≥90 years; age ≥100 years; parental age ≥90th percentile	Parental lifespan; age attained by parents	Yes	Yes	Multiple
CHRNA3 and *CHRNA5*[143,144]	Parental age ≥90th percentile	Parental lifespan; age attained by parents	Yes	No	Cancer
LPA[143,144]	Parental age ≥90th percentile	Parental lifespan; age attained by parents	Yes	No	Multiple
CDKN2A and *CDKN2B*[143]	Parental age ≥90th percentile	Parental lifespan; age attained by parents	Yes	No	Multiple
USP42[141]	Age ≥99th percentile	None	Yes	No	None
TMTC2[141]	Age ≥99th percentile	None	Yes	No	None
IL6[145]	Age ≥100 years	None	No	No	Inflammatory
ANKRD20A9P[145]	Age ≥100 years	None	No	No	None
LINC02227[142]	Age ≥90 years	None	Yes	No	Cardiovascular
FOXO3A[146]	Age ≥90 years	None	Yes	No	None
RAD50 and *IL13*[147]	Age ≥90 years	None	Yes	No	None
MC2R[143]	Parental age ≥90th percentile	None	Yes	No	None
USP2-AS1[143]	Parental age ≥90th percentile	None	Yes	No	None
HLA-DQA1 and *HLA-DRB1*[143,144]	None	Parental lifespan; age attained by parents	Yes	No	Inflammatory
ATXN2[143]	None	Age attained by parents	No	No	Multiple
FURIN[143]	None	Age attained by parents	No	No	Cardiovascular
EPHX2[143]	None	Age attained by parents	No	No	Cancer
PROX2[143]	None	Age attained by parents	No	No	None
CELSR2 and *PSRC1*[143]	None	Age attained by parents	No	No	Cardiovascular

We included only studies that showed one or more genome-wide significant associations with lifespan-related phenotypes ($P < 5 \times 10^{-8}$), with the exception of the *RAD50* and *IL13* locus ($P = 5.42 \times 10^{-7}$), which was based on the number of linkage disequilibrium-independent markers on the genotyping array (Immunochip) used in the study[147]. We excluded studies that were based on results from cohorts that were also included in more recent and larger studies. 'Within publication' refers to replication of a locus in different cohorts within the same publication. 'Between publications' refers to replication of a locus in different cohorts from different publications.

and secretion of bioactive molecules, the senescence-associated secretory phenotype (SASP). Cellular senescence is important during both development[116] and wound healing[117], where it has a key role in tissue remodelling, but in these contexts the senescent cells are eventually removed by macrophages. During ageing, senescent cells persist. Their presence can cause tissue damage, and they are implicated in the aetiology of human age-related diseases, including atherosclerosis, osteoarthritis and cancer[78,118,119]. Selective removal of senescent cells, or disruption of the SASP, can restore tissue homeostasis and increase healthspan and lifespan in mice[118–121]. Although more work in animals will be needed to assess the long-term effects and side effects of this type of intervention, research is already directed towards the possibility of improving the quality of tissues for transplantation, such as kidneys, by prior removal of senescent cells[122] and clinical trials are underway for the treatment of osteoarthritis and glaucoma. A promising approach that has emerged from work on animals is epigenetic reprogramming of aged cells to rejuvenate tissues[123], which has extended lifespan in a mouse model of premature ageing[124]. The myriad of microorganisms present in the gut, the 'microbiome' is increasingly implicated in the health of the gut itself and of other organs during ageing[125,126]. Although most work thus far has been descriptive rather than experimental[127], transfer of the microbiome from young to middle-aged turquoise killifish resulted in an increase in lifespan and a delay in behavioural decline relative to fish that received a transfer from middle-aged fish[128]. The composition of the human gut microbiome shows marked individual variation and is sensitive to many environmental factors, including habitual diet, medication and long-term residential care[129]. Faecal transplantation from lean donors to patients with metabolic syndromes can improve insulin sensitivity[127,130] and probiotic studies in humans are underway, following positive results in mice and safety assessment in humans[131]. Further experimental studies in animals are needed to explore the role of the microbiome in ageing and age-related disease, and to use the findings to inform the design of trials in humans. The systemic, circulatory environment has also proved to play a key part in ageing. Experiments in which the blood systems of mice were conjoined (parabiosis) showed that impaired function of stem cells in multiple aged tissues could be slowed or even reversed[132]. Transfer of blood or plasma, and of plasma proteins, from human umbilical cords has recently been shown to rejuvenate hippocampal function in old mice[132], suggesting that there may be evolutionary conservation of the effector molecules between mice and humans. Identification of these is a high priority for research. The practical accessibility of both the human microbiome and blood system makes therapeutic manipulation a particularly attractive approach, but research in animals is needed to establish the long-term consequences and possible side effects.

Integrating research in animals and humans

The increasing pace of discovery of the mechanisms of ageing in animals, burgeoning practical efforts to characterize and predict the phenotypes of human ageing, together with the recent appearance of databases of electronic health records[133], biobanks and more focused long-term cohort studies, are opening new opportunities to discover the mechanisms that underlie the diversity in physiological deterioration, multimorbidity and frailty and to intervene so we can attenuate or prevent these age-related problems. Further progress will be facilitated by collaboration between scientists who work in different fields. This will align efforts to test the effects of feasible interventions in humans and animal models on ageing biomarkers, hallmarks, multimorbidity and frailty at the individual level. Direct and standardized measures of

Table 2 | Hallmarks of ageing investigated in human studies

Hallmark	Description	Change with age/health (observational studies)	Beneficial response (intervention studies)	Causal evidence
Genomic instability[22,77,148]	Accumulation of genetic damage affects DNA integrity and stability	Accumulation of somatic nuclear and mitochondrial mutations	Dietary energy restriction and increased exercise reduce oxidative and DNA damage, zinc de/repletion and DNA single-stranded breaks	Mutations cause premature ageing syndromes
		Pathogenesis of cancer and progeroid syndromes		
Telomere attrition[22,148,149]	Chromosome caps formed by repeated DNA shortening of telomeric DNA and DNA damage to telomeres	Shortening of telomeres in cells and tissues owing to cell division and damage associated with organ failure, disease and mortality	Mediterranean and plant-based diet and anti-oxidant supplementation slow down telomere shortening, CVD and mortality	Mutations in telomerase cause familiar premature disease, pulmonary fibrosis, dyskeratosis congenita and aplastic anaemia
		Accelerated by smoking and obesity	Bariatric surgery	Loss of regenerative capacity
Epigenetic alterations[22,148,150–152]	Change in DNA methylation, non-coding RNA, histone modification and transcription	Global demethylation and, at promotor regions, hypermethylation and increased variation at polycomb target regions in multiple tissues	Folate and polyphenol supplementation	
		Induced by environmental effects (smoking, stress, trauma, alcohol and sun)	Dietary energy restriction	
		Epigenetic clocks associate with health and disease in prospective studies		
Loss of proteostasis[22,148,153]	Affects protein folding, degradation and repair by ubiquitin proteasome and lysosome autophagy affects synthesis of chaperones	Misfolded and aggregated proteins (in cataracts) and accumulation of autophagic vesicles in affected neurons in neurodegenerative disease	Dietary energy restriction	Mutations in *PS1* and *PS2* cause familial autosomal-dominant Alzheimer's disease and result in amyloid deposition, neuronal loss and lysosome pathology
		Accelerated by obesity		
		Autophagy is better maintained in cente narians and their families	(Intermittent) fasting	
Deregulated nutrient sensing[22,95,148,154,155]	Detect concentrations of intra/extracellular nutrients (glucose, amino acids, AMP, NAD^+) by insulin IGF-1 (IIS), expression of mTOR signalling-induced FOXO transcription factors AMPK and sirtuin	Increased gene expression of mTOR and IIS pathways with age in different tissues and with severity of brain disease	Low protein intake in a cohort study and dietary restriction were associated with low serum IGF-1, but not in all studies, and restored insulin sensitivity	Mutations lowering growth hormone and IGF-1 lower the incidence of cancer and CVD
		Serum IGF-1 decreases with age and is associated with sarcopenia		
		IIS gene variants are associated with long life (85+)		
Mitochondrial dysfunction[22,148,156–159]	Decreased numbers with age compromises mitochondrial function upon energy demand, accumulation of reactive oxygen species, lipid peroxidation, impaired clearance of dysfunctional mitochondria (mitophagy)	Accumulation of reactive oxygen species and somatic mutations in mitochondrial DNA, clonal expansion and mosaic respiratory chain deficiency in multiple tissues	Dietary restriction stimulates fatty acid oxidation and lowers oxidative damage	Mitochondrial mutations cause diseases with multiple ageing symptoms
		Decreased synthesis of mitochondrial proteins in muscle	High-intensity aerobic interval training in young and old improved cardiorespiratory fitness, muscle mass, protein abundance and insulin sensitivity	
		Diversity of cancers, chronic obstructive pulmonary disease (respiratory disease), atherosclerosis and hypertension	Resveratrol supplementation tested for protection of lungs, cardiovascular and respiratory pathways; inconclusive owing to variability in studies and doses	

Continued

Hallmark	Description	Change with age/health (observational studies)	Beneficial response (intervention studies)	Causal evidence
Cellular senescence[22,78,148,160]	Arrest of cell cycle excretion of proteins (SASP) adversely impacts tissues and affects clearance by inflammasome	Accumulation with age in a variety of tissues preceding disease, but controversies exists whether accumulation occurs in healthy individuals	Senotherapy (clearance of senescent cells) in human cell models in which senescence is induced	Germ line and somatic mutations in *CDKN2A* contributes to increased risk of range of cancers
		Accumulation in pathology (lung, kidney and cartilage), in biopsies and after therapeutic damage	Prevention of accumulation of senescence by metformin in human cell models in which senescence is induced	Senolytic drug treatment of human osteoarthritic cartilage explants and cultures: depletion of senescent cells, chondrocyte proliferation and growth of the extracellular matrix in cartilage
		Association of genetic variation at the *CDKN2A–CDKN2B* locus and multiple metabolic diseases	Compounds inducing senescence tested in cancer cells	
Stem-cell exhaustion[161–164]	Decrease in the regenerative potential of stem cells	Observed in pulmonary fibrosis	Regenerative medicine on the basis of mesenchymal stem cells, musculoskeletal damage repair	
		Loss of satellite cells in muscle and decreased regeneration capacity		
		Increased frequency of haematopoietic stem cells with impaired functionality and clonal expansion; however, the health consequences of these impairments remain unclear		
Altered intercellular communication[79]	Deregulated endocrine, neuroendocrine, neuronal signalling associated with chronic inflammation during ageing and decline of adaptive immune system or other inter-organ coordination (such as by the gut microbiome) through blood-borne factors	Chronic inflammation and composition of the gut microbiome	Gastric bypass	
		Chronic overexpression of basal levels of stress-related proteins, such as heat-shock proteins in older patients, ER chaperones, hypoxia factor (HIF1α)	Calorie restriction	
		Poor corresponding adaptive response to stress	Resistance exercise training	

Hallmarks of ageing as formulated for animal studies with adapted criteria: (1) manifestation during normal ageing in cross-sectional (comparison of young and old donors) or longitudinal (repeated measurements over time) studies; (2) aggravation is associated with a pathological condition (accumulates in diseased tissue, prevalent in patients or prospectively predicts health deficit); (3) intervention studies beneficially change aggravation; (4) removal of age-related changes increases health conditions, or aggravation causes accelerated ageing. There is no systematic approach yet to record the hallmarks of ageing in human studies for any of the above criteria and especially repeated measurements in longitudinal studies are missing. Hallmarks may not completely cover all relevant observations in humans, such as the adaptive homeostatic response[79]. Evidence for the causality of the hallmark in human ageing mostly results from mutations causing juvenile forms of age-related disease or ex vivo experimental data, mostly in cell models and sometimes in tissues. *PS1* and *PS2* are also known as *PSEN1* and *PSEN2*, respectively. CVD, cardiovascular disease.

end-life multimorbidity itself are needed, in both animals and humans. Measures of healthspan and of age-specific multimorbidity, although informative, do not directly assess the duration or extent of multimorbidity at the end of life. Few such studies are conducted, because they necessitate longitudinal information on individuals until they die, but they will be necessary to assess the effects of interventions on the compression of morbidity.

The results of research into ageing in animals and humans are producing major dividends. Global public health efforts to increase human healthspan will increasingly focus on lowering the risk of obesity, smoking, high alcohol intake, physical inactivity, hypertension and low-density lipoprotein cholesterol, and success in doing so should yield widespread reductions in diabetes, cardiovascular disease and cancer. Repurposed drugs are also a promising approach to maintain human health during ageing, and new clinical trials are underway with candidates that include mTOR inhibitors[110] and metformin[106]. Drugs that kill senescent cells (senolytics) or that block the SASP also show great promise to induce repair of damaged tissue. If successful, the trials for the treatment of osteoarthritis and glaucoma could be extended to primary prevention among at-risk, elderly people if a consensus can be reached on surrogate end points of cartilage degradation and retinopathy. Ideally, preventative drug treatment in humans would start later in life, to minimize the duration of possible side effects of long-term medication use. However, clinical trials do not, in general, include older people, and evidence that drugs are effective, at which doses and whether they have the expected profile of side effects among the elderly, is needed but is often lacking. For instance, levothyroxine, which is widely used to treat older adults with slightly underactive thyroid function, has proven to be ineffective in older people[134]. The mechanisms leading to this lack of efficacy in late life could be investigated in laboratory animals, particularly to understand whether treatment is effective only if started in middle age or even earlier. Polypharmacy is a major problem in older people, and model organisms could be used to find ways of minimizing its effects. Therapies based on cellular

Box 2

Biomarkers of the physiological state and biological age of individuals

Biomarkers in human research are, on the one hand, used to detect individual variability in the progress of ageing, as risk indicators, and, on the other hand, for monitoring the response to interventions. Different biomarkers have been developed to answer different questions, for example, to monitor the physiological state of individuals, predict the onset and/or progression of age-related diseases, detect the physiological vulnerability of elderly to poor clinical outcomes or predict mortality. Biomarkers of the risk of age-related diseases have been developed with great success. No consensus has yet been reached on biomarkers of biological age, that is, the mismatch between chronological age and the stage of an individual along the ageing process. These biomarkers should ideally meet a number of criteria, such as those defined by the American Federation for Aging Research (AFAR): they should (1) mark the individual stage of ageing and predict mortality better than chronological age; (2) monitor ageing in a range of systems and not the effects of disease; and (3) allow longitudinal tracking (for example, by blood tests or imaging techniques) in animals and humans[171].

Several types of biomarker of the physiological state include whole-system indicators of physical or mental capability (for example, locomotor function, strength, balance, cognition and activity during daily living), physiological reserve (for example, respiratory and cardiovascular function) and the systemic capacities to regulate lipid and glucose metabolism and immunity (for example, insulin, IL-6 and CRP)[33,172]. In addition to single markers, multi-marker indicators have been generated based on assays of multi-organ functionality and/or molecular characteristics. Physiological vulnerability later in life, that is, 'frailty' at ages above 80 years is generally described by low physical activity, muscle weakness, slowed performance, fatigue or poor endurance and unintentional weight loss. About 50 different frailty algorithms are available, the 'frailty phenotype'[173] and 'frailty index'[174] being the most commonly used clinically. For early phases of life, other scores, such as the 'Pace of Aging' score[43], have been generated. More recently, multi-marker indicators of biological age have been based on age-related changes in the transcriptome[175], epigenome[176,177], metabolome[178] and structural neuroimaging[179]. These await systematic testing and comparison with each other and with traditional parameters, in relation to clinical decisions and intervention studies. Different indicators of biological age (telomere shortening, epigenetic clocks and pace of ageing) seem to reflect different aspects of physiological decline[180]. Because long-lasting cohort studies contain many ageing phenotypes and a large amount of clinical, imaging and molecular data, collected at multiple time points, these studies could allow systematic comparisons and development of a multivariate mix of marker profiles with the strongest predictive power.

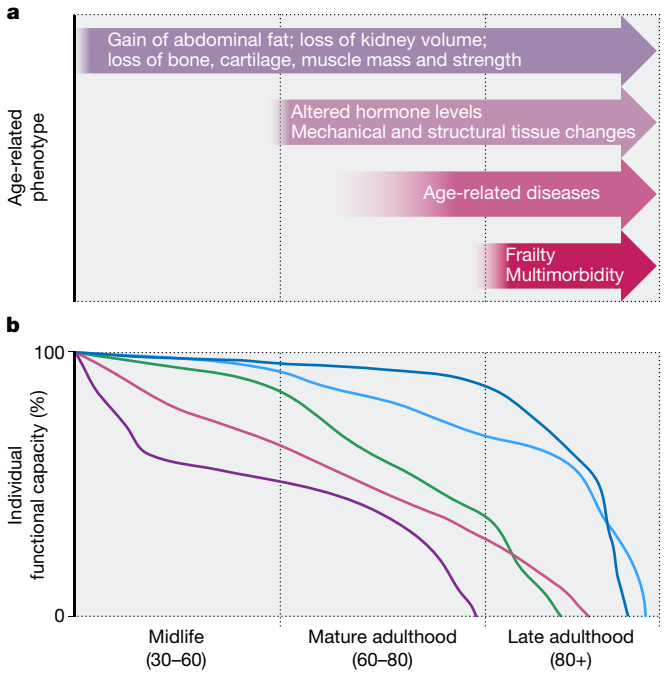


Fig. 3 | Schematic representation of the timing and progression of age-related phenotypes in adult humans. a, Age-related phenotypes include loss of bone and muscle mass, gain of abdominal fat, mechanical and structural tissue changes, age-related diseases and frailty. **b**, Each organ, tissue, cell or trait deteriorates over time at different rates in different people, resulting in individual trajectories of functional decline during ageing, graphically represented by the different coloured lines. The purple line could, for example, represent an individual who rapidly gained abdominal fat during adulthood, reaching a plateau in midlife, with loss of muscle mass and strength in mature adulthood, resulting in a rapid decline of functional capacity of the locomotor system and development of age-related disease, such as osteoarthritis, accompanied by falls and fractures. The dark blue line, on the other hand, could represent someone who remains metabolically healthy until late adulthood, after which he or she suffers from a decline in kidney function, which also affects the cardiovascular system and can result in heart disease, as well as suffering a decline in cognitive capacity and ultimately frailty.

reprogramming and systemic factors from young plasma also show great promise for application in tissue regeneration.

For interventions into the ageing process to have maximum impact on ageing populations, they would ideally be effective as population-wide public-health measures. These would require an excellent safety profile and near-universal efficacy. However, the marked individual variation in the ageing process means that some interventions will be most effective when they are targeted at those people who are most at risk. When establishing risk of rapid physiological decline and age-related disease, and monitoring the response to interventions, blood is the most practically accessible and therefore the most commonly investigated tissue, but it is much less commonly used in animal studies. It will be important to develop blood-based biomarkers of risk, ageing hallmarks and responses to candidate interventions in animals. Mice are commonly used in studies of ageing and age-related disease, but other mammalian species may be more suitable for work on specific conditions, such as rats for thyroid function and blood pressure. Most laboratory mice are also inbred, with marked strain peculiarities, and animals that are more outbred would more closely mirror the individual heterogeneity that is typical of human populations, although this problem is not confined to work on ageing. Some promising new models are also appearing that allow for parallel cell biological studies of animal and human ageing. Direct reprogramming of primary fibroblasts from individuals of different ages can maintain age-specific transcriptional profiles and decreases in nucleocytoplasmic compartmentalization, potentially providing opportunities to study age-related cellular changes in vitro[135]. Organoids can also provide a three-dimensional context for the study of interactions of different cell types with each other and with the extracellular matrix[136]. These systems will facilitate ex vivo work on human ageing with more realistic material than conventional tissue culture.

Further understanding of human ageing is coming from analysis of, for example, electronic health records and biobanks and detailed genetic and phenotypic data from clinical and longitudinal cohort studies. These can capture those features of human ageing that are not recapitulated by laboratory animals. The patterns of age-related disease

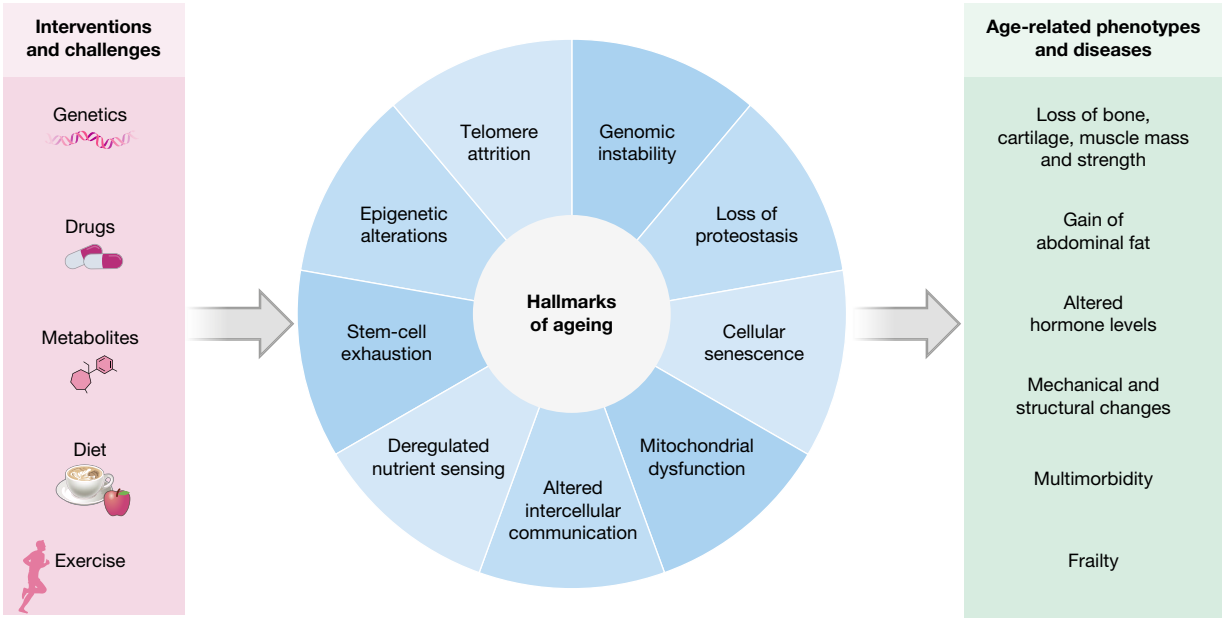


Fig. 4 | Ageing is characterized by mechanistic hallmarks that contribute to ageing to different extents in different organisms, and in different cell types within an organism. Hallmarks can influence each other both within cells and at a distance. Different interventions to prevent or ameliorate symptoms of ageing can affect different groups of hallmarks, and different groups of hallmarks can contribute to the aetiology of specific age-related phenotypes and diseases.

and multimorbidity identified from these sources can then be tested for their association with genetic, molecular and other phenotypic characteristics. Expression of genes, proteins and metabolites associated with age-related diseases can provide more mechanistic insights, including the role of the hallmarks of ageing (Fig. 4). An initial correlation between, for example, increased levels of a protein and the incidence of a health condition can be investigated for causality by Mendelian randomization[137,138], in which the random assignment of genetic variation to individuals at the zygote stage constitutes a natural experiment. Experimental studies in human cells, organoids and animals can then be used to analyse the mechanistic links between the protein and the condition. Data resources, such as the druggable genome[139], can be used to determine whether the protein is a potential drug target of approved or novel drugs that could delay or prevent the condition. These approaches would benefit from standardized protocols to obtain biobank samples from older people at general practitioner and hospital visits, in order to obtain a more representative sample of the elderly population than available from current biobank and cohort studies. Novel assays using metabolic imaging now allow non-invasive recording of metabolic health status[140]. The accumulated longitudinal data and biological specimens that have already been collected in cohort studies can also be used to estimate the individual rate of change in specific biomarkers and outcomes. Robust biomarkers emerging from such systematic research can then be used as surrogate end points to indicate whether anti-ageing interventions are likely to have beneficial effects on clinical outcomes.

The expanding proportion of unhealthy elderly people in many populations is indeed a global challenge to society. However, public health measures to reduce the risk of cancer, metabolic and cardiovascular disease can be effective and should be monitored in primary care. The success of any intervention to combat multimorbidity will be limited by the wish of individuals to reduce its effects and hence their compliance with preventative measures. However, for the willing, lifestyle adjustments and preventative drug treatments are already at hand, with a variety of promising new interventions on the near horizon.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, statements of data availability and associated accession codes are available at https://doi.org/10.1038/s41586-018-0457-8.

Received: 28 March 2018; Accepted: 16 July 2018;
Published online 5 September 2018.

1. Oeppen, J. & Vaupel, J. W. Broken limits to life expectancy. *Science* **296**, 1029–1031 (2002).
2. Waite, L. J. *Aging, Health, and Public Policy: Demographic and Economic Perspectives* (Population Council, 2004).
3. Fogel, R. W. & Costa, D. L. A theory of technophysio evolution, with some implications for forecasting population, health care costs, and pension costs. *Demography* **34**, 49–66 (1997).
4. Vaupel, J. W. et al. Biodemographic trajectories of longevity. *Science* **280**, 855–860 (1998).
5. Dong, X., Milholland, B. & Vijg, J. Evidence for a limit to human lifespan. *Nature* **538**, 257–259 (2016).
6. Kontis, V. et al. Future life expectancy in 35 industrialised countries: projections with a Bayesian model ensemble. *Lancet* **389**, 1323–1335 (2017). **Analysis of age-specific death rates in 35 industrialized countries shows that there is a high probability that life expectancy in these countries will continue to increase during the coming decades.**
7. Christensen, K. et al. Physical and cognitive functioning of people older than 90 years: a comparison of two Danish cohorts born 10 years apart. *Lancet* **382**, 1507–1513 (2013).
8. Zeng, Y., Feng, Q., Hesketh, T., Christensen, K. & Vaupel, J. W. Survival, disabilities in activities of daily living, and physical and cognitive functioning among the oldest-old in China: a cohort study. *Lancet* **389**, 1619–1629 (2017).
9. Burger, O., Baudisch, A. & Vaupel, J. W. Human mortality improvement in evolutionary context. *Proc. Natl Acad. Sci. USA* **109**, 18210–18214 (2012).
10. Poulain, M., Herm, A. & Pes, G. The blue zones: areas of exceptional longevity around the world. Vienna Yearb. *Popul. Res.* **11**, 87–108 (2013).
11. Cooper, R., Strand, B. H., Hardy, R., Patel, K. V. & Kuh, D. Physical capability in mid-life and survival over 13 years of follow-up: British birth cohort study. *Br. Med. J.* **348**, g2219 (2014).
12. Crimmins, E. M. Lifespan and healthspan: past, present, and promise. *Gerontologist* **55**, 901–911 (2015).
13. Jagger, C. et al. Inequalities in healthy life years in the 25 countries of the European Union in 2005: a cross-national meta-regression analysis. *Lancet* **372**, 2124–2131 (2008).
14. *World Report on Ageing and Health*. http://who.int/ageing/events/world-report-2015-launch/en/ (WHO, 2015).
15. Stenholm, S. et al. Body mass index as a predictor of healthy and disease-free life expectancy between ages 50 and 75: a multicohort study. *Int. J. Obes.* **41**, 769–775 (2017).
16. Niccoli, T. & Partridge, L. Ageing as a risk factor for disease. *Curr. Biol.* **22**, R741–R752 (2012).
17. Christensen, K., McGue, M., Petersen, I., Jeune, B. & Vaupel, J. W. Exceptional longevity does not result in excessive levels of disability. *Proc. Natl Acad. Sci. USA* **105**, 13274–13279 (2008). **A survey of a Danish cohort born in 1905 and followed for physical and cognitive independence from 1998 to 2005 across the age range of 92–100 years shows that lifespan extension of a population does not necessarily results in exceptional levels of disability at high ages.**

18. Andersen, S. L., Sebastiani, P., Dworkis, D. A., Feldman, L. & Perls, T. T. Health span approximates life span among many supercentenarians: compression of morbidity at the approximate limit of life span. *J. Gerontol. A* **67A**, 395–405 (2012).
19. Kenyon, C. J. The genetics of ageing. *Nature* **464**, 504–512 (2010).
20. Fontana, L. & Partridge, L. Promoting health and longevity through diet: from model organisms to humans. *Cell* **161**, 106–118 (2015).
21. Fontana, L., Partridge, L. & Longo, V. D. Extending healthy life span—from yeast to humans. *Science* **328**, 321–326 (2010).
22. López-Otín, C., Blasco, M. A., Partridge, L., Serrano, M. & Kroemer, G. The hallmarks of aging. *Cell* **153**, 1194–1217 (2013).
23. van den Berg, N., Beekman, M., Smith, K. R., Janssens, A. & Slagboom, P. E. Historical demography and longevity genetics: back to the future. *Ageing Res. Rev.* **38**, 28–39 (2017).
24. Kaplanis, J. et al. Quantitative analysis of population-scale family trees with millions of relatives. *Science* **360**, 171–175 (2018).
25. Hjelmborg, J. et al. Genetic influence on human lifespan and longevity. *Hum. Genet.* **119**, 312–321 (2006).
26. Slagboom, P. E., van den Berg, N. & Deelen, J. Phenome and genome based studies into human ageing and longevity: An overview. *Biochim. Biophys. Acta* https://doi.org/10.1016/j.bbadis.2017.09.017 (2017).
27. Mahley, R. W. & Rall, S. C. Jr. Apolipoprotein E: far more than a lipid transport protein. *Annu. Rev. Genomics Hum. Genet.* **1**, 507–537 (2000).
28. Sebastiani, P., Nussbaum, L., Andersen, S. L., Black, M. J. & Perls, T. T. Increasing sibling relative risk of survival to older and older ages and the importance of precise definitions of "aging," "life span," and "longevity". *J. Gerontol. A* **71**, 340–346 (2016).
29. Nygaard, M. et al. Birth cohort differences in the prevalence of longevity-associated variants in *APOE* and *FOXO3A* in Danish long-lived individuals. *Exp. Gerontol.* **57**, 41–46 (2014).
30. Terry, D. F. et al. Lower all-cause, cardiovascular, and cancer mortality in centenarians' offspring. *J. Am. Geriatr. Soc.* **52**, 2074–2076 (2004).
31. Westendorp, R. G. et al. Nonagenarian siblings and their offspring display lower risk of mortality and morbidity than sporadic nonagenarians: The Leiden Longevity Study. *J. Am. Geriatr. Soc.* **57**, 1634–1637 (2009).
32. Newman, A. B. et al. Health and function of participants in the Long Life Family Study: a comparison with other cohorts. *Aging (Albany NY)* **3**, 63–76 (2011).
33. Deelen, J. et al. Employing biomarkers of healthy ageing for leveraging genetic studies into human longevity. *Exp. Gerontol.* **82**, 166–174 (2016).
34. Ash, A. S. et al. Are members of long-lived families healthier than their equally long-lived peers? Evidence from the Long Life Family Study. *J. Gerontol. A* **70**, 971–976 (2015).
35. Beekman, M. et al. Genome-wide association study (GWAS)-identified disease risk alleles do not compromise human longevity. *Proc. Natl Acad. Sci. USA* **107**, 18046–18049 (2010).
36. Erikson, G. A. et al. Whole-genome sequencing of a healthy aging cohort. *Cell* **165**, 1002–1011 (2016).
37. Bergman, A., Atzmon, G., Ye, K., MacCarthy, T. & Barzilai, N. Buffering mechanisms in aging: a systems approach toward uncovering the genetic component of aging. *PLOS Comput. Biol.* **3**, e170 (2007).
38. Suh, Y. et al. Functionally significant insulin-like growth factor I receptor mutations in centenarians. *Proc. Natl Acad. Sci. USA* **105**, 3438–3442 (2008).
39. Druley, T. E. et al. Candidate gene resequencing to identify rare, pedigree-specific variants influencing healthy aging phenotypes in the long life family study. *BMC Geriatr.* **16**, 80 (2016).
40. Morris, B. J., Willcox, D. C., Donlon, T. A. & Willcox, B. J. *FOXO3*: a major gene for human longevity—a mini-review. *Gerontology* **61**, 515–525 (2015).
41. Flachsbart, F. et al. Association of *FOXO3A* variation with human longevity confirmed in German centenarians. *Proc. Natl Acad. Sci. USA* **106**, 2700–2705 (2009).
A genetic variant in *FOXO3A*, rs2802292, is associated with longevity in humans.
42. Willcox, B. J. et al. *FOXO3A* genotype is strongly associated with human longevity. *Proc. Natl Acad. Sci. USA* **105**, 13987–13992 (2008).
43. Belsky, D. W. et al. Quantification of biological aging in young adults. *Proc. Natl Acad. Sci. USA* **112**, E4104–E4110 (2015).
44. Bektas, A., Schurman, S. H., Sen, R. & Ferrucci, L. Aging, inflammation and the environment. *Exp. Gerontol.* **105**, 10–18 (2018).
45. Chahal, H. S. & Drake, W. M. The endocrine system and ageing. *J. Pathol.* **211**, 173–180 (2007).
46. Lakatta, E. G. & Levy, D. Arterial and cardiac aging: major shareholders in cardiovascular disease enterprises. Part II: the aging heart in health: links to heart disease. *Circulation* **107**, 346–354 (2003).
47. Barnett, K. et al. Epidemiology of multimorbidity and implications for health care, research, and medical education: a cross-sectional study. *Lancet* **380**, 37–43 (2012).
Analysis of Scottish health registry data from 2007 to 2012 shows a high level of multimorbidity (two or more disorders) in those over age 65 and a 10–15 years earlier onset of multimorbidity in people living in socio-economically deprived areas, challenging the single-disease clinical framework and advocating personalized approaches.
48. Crimmins, E. M., Kim, J. K. & Seeman, T. E. Poverty and biological risk: the earlier "aging" of the poor. *J. Gerontol. A* **64A**, 286–292 (2009).
49. McGuigan, F. E., Bartosch, P. & Åkesson, K. E. Musculoskeletal health and frailty. *Best Pract. Res. Clin. Rheumatol.* **31**, 145–159 (2017).
50. Crimmins, E., Kim, J. K. & Vasunilashorn, S. Biodemography: new approaches to understanding trends and differences in population health and mortality. *Demography* **47**, S41–S64 (2010).
51. Ensrud, K. E. et al. Frailty and risk of falls, fracture, and mortality in older women: the study of osteoporotic fractures. *J. Gerontol. A* **62**, 744–751 (2007).
52. Marengoni, A. et al. Aging with multimorbidity: a systematic review of the literature. *Ageing Res. Rev.* **10**, 430–439 (2011).
53. Guthrie, B., Makubate, B., Hernandez-Santiago, V. & Dreischulte, T. The rising tide of polypharmacy and drug–drug interactions: population database analysis 1995–2010. *BMC Med.* **13**, 74 (2015).
54. Gu, Q., Dillon, C. F. & Burt, V. L. Prescription drug use continues to increase: U.S. prescription drug data for 2007–2008. *NCHS Data Brief* **42**, 1–8 (2010).
55. Bushardt, R. L., Massey, E. B., Simpson, T. W., Ariail, J. C. & Simpson, K. N. Polypharmacy: misleading, but manageable. *Clin. Interv. Aging* **3**, 383–389 (2008).
56. Parameswaran Nair, N. et al. Hospitalization in older patients due to adverse drug reactions — the need for a prediction tool. *Clin. Interv. Aging* **11**, 497–505 (2016).
57. Marcum, Z. A. et al. Prevalence of unplanned hospitalizations caused by adverse drug reactions in older veterans. *J. Am. Geriatr. Soc.* **60**, 34–41 (2012).
58. Howard, R. L. et al. Which drugs cause preventable admissions to hospital? A systematic review. *Br. J. Clin. Pharmacol.* **63**, 136–147 (2007).
59. Stringhini, S. et al. Socioeconomic status and the 25 × 25 risk factors as determinants of premature mortality: a multicohort study and meta-analysis of 1·7 million men and women. *Lancet* **389**, 1229–1237 (2017).
A plea to include socio-economic factors into the initiative of high income WHO member states to cut mortality due to non-communicable diseases by 25% by 2025, showing in a very large prospective multi-cohort meta-analysis study the years-of-life-lost due to high alcohol intake, physical inactivity, current smoking, hypertension, diabetes, obesity and low socio-economic status.
60. Harvey, J. A., Chastin, S. F. & Skelton, D. A. How Sedentary are older people? A systematic review of the amount of sedentary behavior. *J. Aging Phys. Act.* **23**, 471–487 (2015).
61. *Global Action Plan for the Prevention and Control of NCDs 2013–2020*. http://who.int/nmh/events/ncd_action_plan/en/ (WHO, 2013).
62. Diabetes Prevention Program Research Group. Long-term effects of lifestyle intervention or metformin on diabetes development and microvascular complications over 15-year follow-up: the Diabetes Prevention Program Outcomes Study. *Lancet Diabetes Endocrinol.* **3**, 866–875 (2015).
63. Estruch, R. et al. Primary prevention of cardiovascular disease with a mediterranean diet supplemented with extra-virgin olive oil or nuts. *N. Engl. J. Med.* **378**, e34 (2018).
64. Toledo, E. et al. Mediterranean diet and invasive breast cancer risk among women at high cardiovascular risk in the PREDIMED trial: a randomized clinical trial. JAMA *Intern. Med.* **175**, 1752–1760 (2015).
65. Penedo, F. J. & Dahn, J. R. Exercise and well-being: a review of mental and physical health benefits associated with physical activity. *Curr. Opin. Psychiatry* **18**, 189–193 (2005).
66. Heilbronn, L. K. et al. Effect of 6-month calorie restriction on biomarkers of longevity, metabolic adaptation, and oxidative stress in overweight individuals: a randomized controlled trial. *J. Am. Med. Assoc.* **295**, 1539–1548 (2006).
67. Ngandu, T. et al. A 2 year multidomain intervention of diet, exercise, cognitive training, and vascular risk monitoring versus control to prevent cognitive decline in at-risk elderly people (FINGER): a randomised controlled trial. *Lancet* **385**, 2255–2263 (2015).
A two-year proof-of-concept randomized clinical trial in the Finnish Geriatric Intervention Study to Prevent Cognitive Impairment and Disability (FINGER), in which positive results are found for a multi-domain approach (diet, exercise, cognitive training and vascular risk monitoring) compared to general health advice to prevent cognitive decline in at-risk elderly people (60–77 years of age) from the general population.
68. Most, J., Tosti, V., Redman, L. M. & Fontana, L. Calorie restriction in humans: an update. *Ageing Res. Rev.* **39**, 36–45 (2017).
69. Mattson, M. P., Longo, V. D. & Harvie, M. Impact of intermittent fasting on health and disease processes. *Ageing Res. Rev.* **39**, 46–58 (2017).
70. Laird, E. et al. The prevalence of vitamin D deficiency and the determinants of 25(OH)D concentration in older Irish adults: data from The Irish Longitudinal Study on Ageing (TILDA). *J. Gerontol. A* **73**, 519–525 (2018).
71. Levine, M. E. et al. Low protein intake is associated with a major reduction in IGF-1, cancer, and overall mortality in the 65 and younger but not older population. *Cell Metab.* **19**, 407–417 (2014).
72. Ettehad, D. et al. Blood pressure lowering for prevention of cardiovascular disease and death: a systematic review and meta-analysis. *Lancet* **387**, 957–967 (2016).
73. Collins, R. et al. Interpretation of the evidence for the efficacy and safety of statin therapy. *Lancet* **388**, 2532–2561 (2016).
74. Bibbins-Domingo, K. et al. Statin use for the primary prevention of cardiovascular disease in adults: US preventive services task force recommendation statement. *J. Am. Med. Assoc.* **316**, 1997–2007 (2016).
75. Ahmadi, S. F. et al. Reverse epidemiology of traditional cardiovascular risk factors in the geriatric population. *J. Am. Med. Dir. Assoc.* **16**, 933–939 (2015).
76. Conroy, S. P., Westendorp, R. G. J. & Witham, M. D. Hypertension treatment for older people—navigating between Scylla and Charybdis. *Age Ageing* (2018).
77. Vijg, J. *Aging of the Genome: The Dual Role of DNA in Life and Death* (Oxford Univ. Press, Oxford, 2007).

78. Yanai, H. & Fraifeld, V. E. The role of cellular senescence in aging through the prism of Koch-like criteria. *Ageing Res. Rev.* **41**, 18–33 (2018).
79. Pomatto, L. C. D. & Davies, K. J. A. The role of declining adaptive homeostasis in ageing. *J. Physiol. (Lond.)* **595**, 7275–7309 (2017).
80. Franceschi, C. & Campisi, J. Chronic inflammation (inflammaging) and its potential contribution to age-associated diseases. *J. Gerontol. A* **69**, S4–S9 (2014).
81. Mitchell, S. J., Scheibye-Knudsen, M., Longo, D. L. & de Cabo, R. Animal models of aging research: implications for human aging and age-related diseases. *Annu. Rev. Anim. Biosci.* **3**, 283–303 (2015).
82. Tiku, V. et al. Small nucleoli are a cellular hallmark of longevity. *Nat. Commun.* **8**, 16083 (2016).
83. Heidinger, B. J. et al. Telomere length in early life predicts lifespan. *Proc. Natl Acad. Sci. USA* **109**, 1743–1748 (2012).
84. Varela, E., Muñoz-Lorente, M. A., Tejera, A. M., Ortega, S. & Blasco, M. A. Generation of mice with longer and better preserved telomeres in the absence of genetic manipulations. *Nat. Commun.* **7**, 11739 (2016).
85. Kapahi, P., Kaeberlein, M. & Hansen, M. Dietary restriction and lifespan: lessons from invertebrate models. *Ageing Res. Rev.* **39**, 3–14 (2017).
86. Colman, R. J. et al. Caloric restriction reduces age-related and all-cause mortality in rhesus monkeys. *Nat. Commun.* **5**, 3557 (2014).
87. Mattison, J. A. et al. Impact of caloric restriction on health and survival in rhesus monkeys from the NIA study. *Nature* **489**, 318–321 (2012).
88. Vaughan, K. L. et al. Caloric restriction study design limitations in rodent and nonhuman primate studies. *J. Gerontol. A* **73**, 48–53 (2018).
89. Pan, H. & Finkel, T. Key proteins and pathways that regulate lifespan. *J. Biol. Chem.* **292**, 6452–6460 (2017).
90. Pawlikowska, L. et al. Association of common genetic variation in the insulin/IGF1 signaling pathway with human longevity. *Aging Cell* **8**, 460–472 (2009).
91. Deelen, J. et al. Gene set analysis of GWAS data for human longevity highlights the relevance of the insulin/IGF-1 signaling and telomere maintenance pathways. *Age (Dordr.)* **35**, 235–249 (2013).
92. Passtoors, W. M. et al. Gene expression analysis of mTOR pathway: association with human longevity. *Aging Cell* **12**, 24–31 (2013).
93. Johnson, S. C., Rabinovitch, P. S. & Kaeberlein, M. mTOR is a key modulator of ageing and age-related disease. *Nature* **493**, 338–345 (2013).
94. Fruman, D. A. et al. The PI3K pathway in human disease. *Cell* **170**, 605–635 (2017).
95. Longo, V. D. et al. Interventions to slow aging in humans: are we ready? *Aging Cell* **14**, 497–510 (2015).
96. Ingram, D. K. & de Cabo, R. Calorie restriction in rodents: caveats to consider. *Ageing Res. Rev.* **39**, 15–28 (2017).
97. Racette, S. B. et al. One year of caloric restriction in humans: feasibility and effects on body composition and abdominal adipose tissue. *J. Gerontol. A* **61**, 943–950 (2006).
98. Solon-Biet, S. M. et al. The ratio of macronutrients, not caloric intake, dictates cardiometabolic health, aging, and longevity in ad libitum-fed mice. *Cell Metab.* **19**, 418–430 (2014).
99. Grandison, R. C., Piper, M. D. & Partridge, L. Amino-acid imbalance explains extension of lifespan by dietary restriction in *Drosophila*. *Nature* **462**, 1061–1064 (2009).
100. Manoogian, E. N. C. & Panda, S. Circadian rhythms, time-restricted feeding, and healthy aging. *Ageing Res. Rev.* **39**, 59–67 (2017).
101. Acosta-Rodriguez, V. A., de Groot, M. H. M., Rijo-Ferreira, F., Green, C. B. & Takahashi, J. S. Mice under caloric restriction self-impose a temporal restriction of food intake as revealed by an automated feeder system. *Cell Metab.* **26**, 267–277 (2017).
102. Xie, K. et al. Every-other-day feeding extends lifespan but fails to delay many symptoms of aging in mice. *Nat. Commun.* **8**, 155 (2017).
103. Fontana, L. The science of nutritional modulation of aging. *Ageing Res. Rev.* **39**, 1–2 (2017).
104. Wei, M. et al. Fasting-mimicking diet and markers/risk factors for aging, diabetes, cancer, and cardiovascular disease. *Sci. Transl. Med.* **9**, eaai8700 (2017).
Healthy human subjects randomly allocated to a diet that mimics fasting, which is low in calories, sugars and protein, but high in unsaturated fats, as opposed to unrestricted food consumption, showed reduced body weight, trunk and total body fat, had lower blood pressure and decreased levels of IGF-1, with more marked effects in participants at risk of disease.
105. Longo, V. D. & Panda, S. Fasting, circadian rhythms, and time-restricted feeding in healthy lifespan. *Cell Metab.* **23**, 1048–1059 (2016).
106. Newman, J. C. et al. Strategies and challenges in clinical trials targeting human aging. *J. Gerontol. A* **71**, 1424–1434 (2016).
107. Blenis, J.TOR, the gateway to cellular metabolism, cell growth, and disease. *Cell* **171**, 10–13 (2017).
108. Slack, C. Ras signaling in aging and metabolic regulation. *Nutr. Healthy Aging* **4**, 195–205 (2017).
109. Bitto, A. et al. Transient rapamycin treatment can increase lifespan and healthspan in middle-aged mice. *eLife* **5**, e16351 (2016).
110. Mannick, J. B. et al. mTOR inhibition improves immune function in the elderly. *Sci. Transl. Med.* **6**, 268ra179 (2014).
At doses that were well-tolerated, the mTOR inhibitor RAD001 enhanced the response to the influenza vaccine in elderly volunteers by about 20% and reduced the percentage of $CD4^+$ and $CD8^+$ T lymphocytes that expressed the programmed death-1 receptor, which inhibits T cell signalling and shows higher expression with increasing age.
111. Strong, R. et al. Longer lifespan in male mice treated with a weakly estrogenic agonist, an antioxidant, an α-glucosidase inhibitor or a Nrf2-inducer. *Aging Cell* **15**, 872–884 (2016).
112. Barzilai, N., Crandall, J. P., Kritchevsky, S. B. & Espeland, M. A. Metformin as a tool to target aging. *Cell Metab.* **23**, 1060–1065 (2016).
113. Espeland, M. A. et al. Clinical trials targeting aging and age-related multimorbidity. *J. Gerontol. A* **72**, 355–361 (2017).
114. Arriola Apelo, S. I., Pumper, C. P., Baar, E. L., Cummings, N. E. & Lamming, D. W. Intermittent administration of rapamycin extends the life span of female C57BL/6J mice. *J. Gerontol. A* **71**, 876–881 (2016).
115. Johnson, S. C. & Kaeberlein, M. Rapamycin in aging and disease: maximizing efficacy while minimizing side effects. *Oncotarget* **7**, 44876–44878 (2016).
116. Muñoz-Espín, D. et al. Programmed cell senescence during mammalian embryonic development. *Cell* **155**, 1104–1118 (2013).
117. Demaria, M. et al. An essential role for senescent cells in optimal wound healing through secretion of PDGF-AA. *Dev. Cell* **31**, 722–733 (2014).
118. Childs, B. G. et al. Senescent cells: an emerging target for diseases of ageing. *Nat. Rev. Drug Discov.* **16**, 718–735 (2017).
119. Jeon, O. H. et al. Local clearance of senescent cells attenuates the development of post-traumatic osteoarthritis and creates a pro-regenerative environment. *Nat. Med.* **23**, 775–781 (2017).
Transection of the anterior cruciate ligament in mice caused accumulation of senescent cells in the articular cartilage and synovium, and selective elimination of these cells or injection of a senolytic molecule attenuated the development of osteoarthritis, reduced pain and increased cartilage development.
120. Childs, B. G., Durik, M., Baker, D. J. & van Deursen, J. M. Cellular senescence in aging and age-related disease: from mechanisms to therapy. *Nat. Med.* **21**, 1424–1435 (2015).
121. McHugh, D. & Gil, J. Senescence and aging: causes, consequences, and therapeutic avenues. *J. Cell Biol.* **217**, 65–77 (2018).
122. van Willigenburg, H., de Keizer, P. L. J. & de Bruin, R. W. F. Cellular senescence as a therapeutic target to improve renal transplantation outcome. *Pharmacol. Res.* **130**, 322–330 (2018).
123. Rando, T. A. & Chang, H. Y. Aging, rejuvenation, and epigenetic reprogramming: resetting the aging clock. *Cell* **148**, 46–57 (2012).
124. Ocampo, A. et al. In vivo amelioration of age-associated hallmarks by partial reprogramming. *Cell* **167**, 1719–1733 (2016).
125. Clark, R. I. & Walker, D. W. Role of gut microbiota in aging-related health decline: insights from invertebrate models. *Cell. Mol. Life Sci.* **75**, 93–101 (2018).
126. Kundu, P., Blacher, E., Elinav, E. & Pettersson, S. Our gut microbiome: the evolving inner self. *Cell* **171**, 1481–1493 (2017).
127. Schmidt, T. S. B., Raes, J. & Bork, P. The Human gut microbiome: from association to modulation. *Cell* **172**, 1198–1215 (2018).
128. Smith, P. et al. Regulation of life span by the gut microbiota in the short-lived African turquoise killifish. *eLife* **6**, e27014 (2017).
Recolonizing the gut of middle-age turquoise killifish with bacteria from young, rather than middle-aged, donors extends lifespan, delays behavioural decline and prevents the changes in the microbiome associated with host ageing.
129. O'Toole, P. W. & Jeffery, I. B. Gut microbiota and aging. *Science* **350**, 1214–1215 (2015).
130. Kootte, R. S. et al. Improvement of insulin sensitivity after lean donor feces in metabolic syndrome is driven by baseline intestinal microbiota composition. *Cell Metab.* **26**, 611–619 (2017).
131. Plovier, H. et al. A purified membrane protein from *Akkermansia muciniphila* or the pasteurized bacterium improves metabolism in obese and diabetic mice. *Nat. Med.* **23**, 107–113 (2017).
132. Castellano, J. M. et al. Human umbilical cord plasma proteins revitalize hippocampal function in aged mice. *Nature* **544**, 488–492 (2017).
Treatment with human umbilical cord plasma revitalizes the hippocampus and improves cognitive function in mice, which is (partially) driven by TIMP2.
133. Casey, J. A., Schwartz, B. S., Stewart, W. F. & Adler, N. E. Using electronic health records for population health research: a review of methods and applications. *Annu. Rev. Public Health* **37**, 61–81 (2016).
134. Stott, D. J. et al. Thyroid hormone therapy for older adults with subclinical hypothyroidism. *N. Engl. J. Med.* **376**, 2534–2544 (2017).
135. Mertens, J. et al. Directly reprogrammed human neurons retain aging-associated transcriptomic signatures and reveal age-related nucleocytoplasmic defects. *Cell Stem Cell* **17**, 705–718 (2015).
Direct conversion of human fibroblasts into induced neurons shows that these cells retain their age-related transcriptional profiles, and demonstrates the potential of direct reprogramming for in vitro modelling of ageing.
136. Hu, J. L., Todhunter, M. E., LaBarge, M. A. & Gartner, Z. J. Opportunities for organoids as new models of aging. *J. Cell Biol.* **217**, 39–50 (2018).
137. Swerdlow, D. I. et al. Selecting instruments for Mendelian randomization in the wake of genome-wide association studies. *Int. J. Epidemiol.* **45**, 1600–1616 (2016).
138. Ainsworth, H. F., Shin, S. Y. & Cordell, H. J. A comparison of methods for inferring causal relationships between genotype and phenotype using additional biological measurements. *Genet. Epidemiol.* **41**, 577–586 (2017).
139. Finan, C. et al. The druggable genome and support for target identification and validation in drug development. *Sci. Transl. Med.* **9**, eaag1166 (2017).
140. Schrauwen-Hinderling, V. B. & Schols, A. M. W. J. Imaging in metabolic research: challenges and opportunities. *J. Appl. Physiol.* **124**, 160–161 (2018).

141. Sebastiani, P. et al. Four genome-wide association studies identify new extreme longevity variants. *J. Gerontol. A* **72**, 1453–1464 (2017).
142. Deelen, J. et al. Genome-wide association meta-analysis of human longevity identifies a novel locus conferring survival beyond 90 years of age. *Hum. Mol. Genet.* **23**, 4420–4432 (2014).
143. Pilling, L. C. et al. Human longevity: 25 genetic loci associated in 389,166 UK biobank participants. *Aging (Albany NY)* **9**, 2504–2520 (2017).
144. Joshi, P. K. et al. Genome-wide meta-analysis associates *HLA-DQA1/DRB1* and *LPA* and lifestyle factors with human longevity. *Nat. Commun.* **8**, 910 (2017).
145. Zeng, Y. et al. Novel loci and pathways significantly associated with longevity. *Sci. Rep.* **6**, 21243 (2016).
146. Broer, L. et al. GWAS of longevity in CHARGE consortium confirms *APOE* and *FOXO3* candidacy. *J. Gerontol. A* **70**, 110–118 (2015).
147. Flachsbart, F. et al. Immunochip analysis identifies association of the *RAD50/IL13* region with human longevity. *Aging Cell* **15**, 585–588 (2016).
148. Anderson, R., Richardson, G. D. & Passos, J. F. Mechanisms driving the ageing heart. *Exp. Gerontol.* https://doi.org/10.1016/j.exger.2017.10.015 (2017).
149. Meiners, S., Eickelberg, O. & Königshoff, M. Hallmarks of the ageing lung. *Eur. Respir. J.* **45**, 807–827 (2015).
150. van Dongen, J. et al. Genetic and environmental influences interact with age and sex in shaping the human methylome. *Nat. Commun.* **7**, 11115 (2016).
151. Slieker, R. C. et al. Age-related accrual of methylomic variability is linked to fundamental ageing mechanisms. *Genome Biol.* **17**, 191 (2016).
152. Declerck, K. & Vanden Berghe, W. Back to the future: epigenetic clock plasticity towards healthy aging. *Mech. Ageing Dev.* https://doi.org/10.1016/j.mad.2018.01.002 (2018).
153. Frake, R. A., Ricketts, T., Menzies, F. M. & Rubinsztein, D. C. Autophagy and neurodegeneration. *J. Clin. Invest.* **125**, 65–74 (2015).
154. Sharples, A. P. et al. Longevity and skeletal muscle mass: the role of IGF signalling, the sirtuins, dietary restriction and protein intake. *Aging Cell* **14**, 511–523 (2015).
155. Wahl, D. et al. Nutritional strategies to optimise cognitive function in the aging brain. *Ageing Res. Rev.* **31**, 80–92 (2016).
156. Sterky, F. H., Lee, S., Wibom, R., Olson, L. & Larsson, N. G. Impaired mitochondrial transport and Parkin-independent degeneration of respiratory chain-deficient dopamine neurons in vivo. *Proc. Natl Acad. Sci. USA* **108**, 12937–12942 (2011).
157. Timmers, S. et al. Calorie restriction-like effects of 30 days of resveratrol supplementation on energy metabolism and metabolic profile in obese humans. *Cell Metab.* **14**, 612–622 (2011).
158. Beijers, R. J. H. C. G., Gosker, H. R. & Schols, A. M. W. J. Resveratrol for patients with chronic obstructive pulmonary disease: hype or hope? *Curr. Opin. Clin. Nutr. Metab. Care* **21**, 138–144 (2018).
159. Robinson, M. M. et al. Enhanced protein translation underlies improved metabolic and physical adaptations to different exercise training modes in young and old humans. *Cell Metab.* **25**, 581–592 (2017).
160. Sturmlechner, I., Durik, M., Sieben, C. J., Baker, D. J. & van Deursen, J. M. Cellular senescence in renal ageing and disease. *Nat. Rev. Nephrol.* **13**, 77–89 (2017).
161. Goodell, M. A. & Rando, T. A. Stem cells and healthy aging. *Science* **350**, 1199–1204 (2015).
162. de Haan, G. & Lazare, S. S. Aging of hematopoietic stem cells. *Blood* **131**, 479–487 (2018).
163. Čamernik, K. et al. Mesenchymal stem cells in the musculoskeletal system: from animal models to human tissue regeneration? *Stem Cell Rev.* **14**, 346–369 (2018).
164. van den Akker, E. B. et al. Uncompromised 10-year survival of oldest old carrying somatic mutations in *DNMT3A* and *TET2*. *Blood* **127**, 1512–1515 (2016).
165. WHO. *Disease Burden and Mortality Estimates: Disease Burden*, 2000–2016 http://www.who.int/healthinfo/global_burden_disease/estimates/en/index1.html (WHO, 2018).
166. Walter, S. et al. A genome-wide association study of aging. *Neurobiol. Aging* **32**, 2109.e15–2109.e28 (2011).
167. Schächter, F. et al. Genetic associations with human longevity at the *APOE* and *ACE* loci. *Nat. Genet.* **6**, 29–32 (1994).
Genetic variants in ApoE are associated with longevity in humans, one of these variants, the ApoE ε4 allele, shows a deleterious effect whereas the other, the ApoE ε2 allele, has a protective effect.
168. Budovsky, A. et al. LongevityMap: a database of human genetic variants associated with longevity. *Trends Genet.* **29**, 559–560 (2013).
169. Jeck, W. R., Siebold, A. P. & Sharpless, N. E. Review: a meta-analysis of GWAS and age-associated diseases. *Aging Cell* **11**, 727–731 (2012).
170. Tazearslan, C., Huang, J., Barzilai, N. & Suh, Y. Impaired IGF1R signaling in cells expressing longevity-associated human IGF1R alleles. *Aging Cell* **10**, 551–554 (2011).
171. Johnson, T. E. Recent results: biomarkers of aging. *Exp. Gerontol.* **41**, 1243–1246 (2006).
172. Lara, J. et al. A proposed panel of biomarkers of healthy ageing. *BMC Med.* **13**, 222 (2015).
173. Fried, L. P. et al. Frailty in older adults: evidence for a phenotype. *J. Gerontol. A* **56**, M146–M157 (2001).
The Fried criteria were among the earliest multi-domain definitions to index frailty, including shrinking (weight loss), muscle weakness (handgrip strength), poor endurance (self-reported exhaustion), slowness (gait speed) and low physical activity (kcal expended per week), estimated in community-dwelling people over 65 years, and this definition of frailty is now widely used in large epidemiological and clinical studies and formed the basis for the development of novel indexes to predict elderly people who have a higher risk of incidences of disease, hospitalization, falls, disability and mortality.
174. Mitnitski, A. B., Mogilner, A. J. & Rockwood, K. Accumulation of deficits as a proxy measure of aging. *Sci. World J.* **1**, 323–336 (2001).
175. Peters, M. J. et al. The transcriptional landscape of age in human peripheral blood. *Nat. Commun.* **6**, 8570 (2015).
176. Horvath, S. DNA methylation age of human tissues and cell types. *Genome Biol.* **14**, R115 (2013).
A combination of CpG methylation sites (epigenetic clock) is associated with chronological ageing of multiple tissues and cell types and can be used to estimate the biological age of a person.
177. Hannum, G. et al. Genome-wide methylation profiles reveal quantitative views of human aging rates. *Mol. Cell* **49**, 359–367 (2013).
178. Hertel, J. et al. Measuring biological age via metabonomics: the metabolic age score. *J. Proteome Res.* **15**, 400–410 (2016).
179. Cole, J. H. et al. Brain age predicts mortality. *Mol. Psychiatry* **23**, 1385–1392 (2018).
180. Jylhävä, J., Pedersen, N. L. & Hägg, S. Biological age predictors. *EBioMedicine* **21**, 29–36 (2017).

Acknowledgements We thank N. van den Berg for help with the preparation of Fig. 1 and N. Chaturvedi and B. J. Zwaan for their critical reading of the manuscript. L.P. acknowledges support from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (no. 741989) and a Wellcome Trust Strategic Award. J.D. acknowledges support from the Alexander von Humboldt Foundation. We apologize to the authors of many relevant studies for not citing their work owing to space limitations.

Reviewer information *Nature* thanks V. D. Longo and J. Vijg for their contribution to the peer review of this work.

Author contributions All authors contributed to the design and writing of the Review.

Competing interests The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41586-018-0457-8.
Reprints and permissions information is available at http://www.nature.com/reprints.
Correspondence and requests for materials should be addressed to L.P. or P.E.S.

ARTICLE

https://doi.org/10.1038/s41586-018-0459-6

Integrating time from experience in the lateral entorhinal cortex

Albert Tsao[1,3]*, Jørgen Sugar[1], Li Lu[1,4], Cheng Wang[2], James J. Knierim[2], May-Britt Moser[1] & Edvard I. Moser[1]*

The encoding of time and its binding to events are crucial for episodic memory, but how these processes are carried out in hippocampal–entorhinal circuits is unclear. Here we show in freely foraging rats that temporal information is robustly encoded across time scales from seconds to hours within the overall population state of the lateral entorhinal cortex. Similarly pronounced encoding of time was not present in the medial entorhinal cortex or in hippocampal areas CA3–CA1. When animals' experiences were constrained by behavioural tasks to become similar across repeated trials, the encoding of temporal flow across trials was reduced, whereas the encoding of time relative to the start of trials was improved. The findings suggest that populations of lateral entorhinal cortex neurons represent time inherently through the encoding of experience. This representation of episodic time may be integrated with spatial inputs from the medial entorhinal cortex in the hippocampus, allowing the hippocampus to store a unified representation of what, where and when.

The representation of time is a crucial component of episodic memory[1–3]. Although a considerable body of work has now demonstrated that the hippocampus has an essential role in generating a representation of time[4–11], our understanding of how the brain represents time for episodic memory (episodic time) is still in a nascent stage. The primary function of episodic time is to record the order of events within experience, which does not require a precise representation of metric time, differentiating it from interval and circadian timing[12–14]. Rather than being able to keep precise metric time, the neural code for episodic time should have the following two fundamental properties: 1) the code should arise automatically without any behavioural training, to support one-shot formation of episodic memory, and 2) the code should be able to capture the different scales of time at which experience may occur. Recently, two types of representation of time have been observed in the hippocampus and medial entorhinal cortex (MEC): time cells, which fire at specific points in time as an animal performs a task[15–19], and the decorrelation of place cell activity across hours to days[20–25]. However, neither of these representations of time has been shown to fully support one-shot formation of episodic memories in combination with variable timescales. Furthermore, how either of these representations of time arises is unknown. Here, we investigated temporal coding outside the place-cell system, in the lateral entorhinal cortex (LEC). We focused on the LEC because (i) this area is a major source of cortical input to the hippocampus, (ii) previous work has shown that responses in the LEC to physical stimuli could be unstable across time[26,27], and (iii) a clear underlying function has not yet been defined for the LEC. We found a representation of time in the LEC that exhibited both of the expected signatures of episodic time, and thus could support episodic memory.

Temporal coding in individual LEC cells

To explore temporal coding in the LEC, we recorded neural activity over more than an hour while rats ran in a box in which the colour of the walls alternated between black and white in a fixed pattern over 12 trials (BW12 experiment, Fig. 1a, Extended Data Fig. 1). An extended number of trials was used to increase the likelihood that animals defined multiple temporal contexts across the experiment, and an interleaved design was chosen to avoid confounding changes in wall colour with progression of time. Data were also recorded from the CA3 and MEC for comparison. Examining LEC responses by eye, we noticed that some cells exhibited clear ramping activity (Fig. 1a, b), raising the possibility that the passage of time can be tracked through the firing rates of individual LEC cells. Responses to specific environmental features such as walls and cue cards[26] were also observed, consistent with the established role of the LEC in encoding environmental context[28–30].

We quantified the influence of wall colour and time on the activity of single cells using a generalized linear model (GLM) incorporating time, wall colour and position as variables for fitting the firing rates of individual neurons, which were binned temporally into blocks of 500 ms (Extended Data Fig. 2a–d). A considerable number of LEC cells were selective specifically for time (20.4% of all recorded cells), whereas only 2.0% of CA3 cells and 4.5% of MEC cells were selective for time alone (number of cells significantly influenced by at least one variable for LEC: 186 out of 451, 41.2%; CA3: 72 out of 148, 48.6%; MEC: 49 out of 133, 36.8%; Fig. 1c). The distributions of cell selectivity for the LEC, CA3 and MEC were consistent across individual animals (Fig. 1c).

Because time was modelled as a linearly increasing function in the GLM, all identified time-selective cells exhibited some form of ramping activity. Both increasing and decreasing ramps across a range of time constants were observed (Extended Data Fig. 2c), as would be expected if they performed a Laplace transform of the recent past[31]. Ramping cells were found in both deep and superficial layers, with no clear difference in the time constants (see Supplementary Information). Ramping responses, particularly across the entire recording session, were not due to recording instability (Extended Data Fig. 2e–g, see Methods). We found little evidence in the LEC for non-ramping time-specific activity similar to that of time cells (Extended Data Fig. 3).

LEC population states encode temporal information

We next focused on the overall population of LEC cells and asked whether its dynamics reflected time coding. We first visualized the overall population data from individual animals using linear discriminant analysis to determine the dimensions that optimally discriminate the

[1]Kavli Institute for Systems Neuroscience and Centre for Neural Computation, NTNU, Trondheim, Norway. [2]Zanvyl Krieger Mind/Brain Institute, Johns Hopkins University, Baltimore, MD, USA. [3]Present address: Department of Biology, Stanford University, Stanford, CA, USA. [4]Present address: Department of Neuroscience, Baylor College of Medicine, Houston, TX, USA. *e-mail: albert.tsao@ntnu.no; edvard.moser@ntnu.no

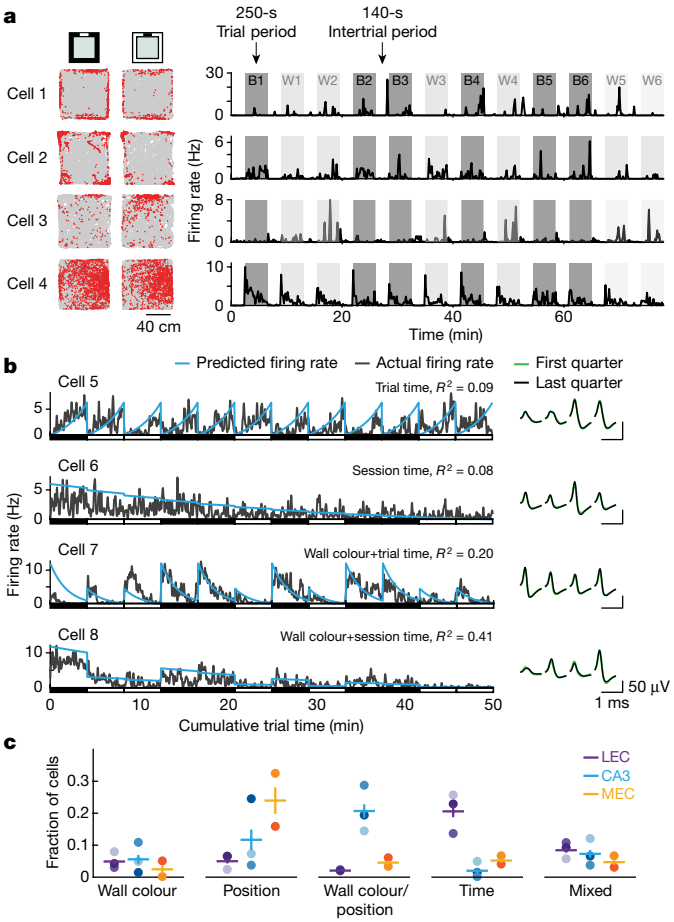


Fig. 1 | Temporal information within LEC single-cell activity. **a**–**c**, Experimental design: animals ran 12 250-s trials in a box with either black or white walls. Trials were separated by 140-s intertrial periods in which the animal was placed in a holding pot. The total session length was 1.3 h. **a**, Four example LEC cells. Left, path plots showing the animal's location in grey, and cell spikes in red. Right, firing rate plots for the same cells. Dark and light grey indicate trial periods with black (B) and white (W) walls, respectively. Unshaded regions indicate the intertrial periods. Cells 1–3 responded to features of the recording box with no obvious temporal component to their activity, whereas cell 4 exhibited ramping activity across each trial. **b**, Left, example GLM fit results for four cells with selectivity for different features, with the observed firing rate shown in grey, and predicted firing rate in blue. Right, average tetrode waveforms for the first (green) and last (black) quarter of the session. **c**, Distribution of selectivity for wall colour, position, time and mixtures of variables for the LEC (top, $n = 3$ animals with 90, 141 and 220 cells), CA3 (middle, $n = 3$ animals with 78, 42 and 28 cells) and MEC (bottom, $n = 2$ animals with 31 and 102 cells). Circles indicate individual animals, solid lines indicate mean ± s.e.m., shading indicates the same individual across the different variables.

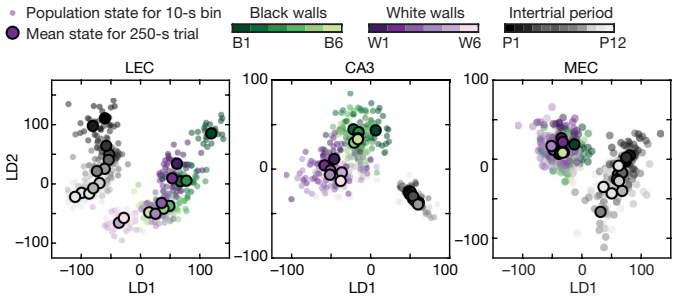


Fig. 2 | Visualizing LEC population activity. 2D projections of neural population responses. Axes correspond to the first two linear discriminants (LD1 and LD2; arbitrary units). Left column shows LEC population responses, middle column shows CA3 population responses, right column shows MEC population responses, each from an example animal. Each trial's wall colour is indicated by a shade of green (black walls) or purple (white walls); intertrial periods are shown in grey. Progression of shade from dark to light for trial and intertrial periods indicates the progression of time. Note progression by time for the LEC (see also Extended Data Fig. 4a). Linear discriminant analysis projections were used for visualization purposes only.

24 experiment-defined states (12 trials, 12 intertrials), and plotting the 2D projection that yielded the best discrimination (Fig. 2, Extended Data Fig. 4a–c). Population activity for the entire session separated into distinct modes corresponding to the environmental contexts of the experiment, as expected from previous work demonstrating LEC involvement in encoding context[28–30]. In addition, there was a prominent progression of states corresponding to the temporal order of the experiment. Thus, across the entire session, the three environmental contexts could be separated along one axis in state space, while the temporal epoch of each trial could be separated along another.

To quantify the temporal information present, we trained a linear multiclass support vector machine to identify temporal epochs (trial 1, intertrial 1, trial 2, and so on) based on neural activity from individual animals, pooled across recording sessions. Population activity defined by the firing rates of cells was binned into 10 s bins, with each bin labelled by the temporal epoch it was in. The decoder was then trained to identify temporal epochs based on population activity, with tenfold cross-validation (Extended Data Fig. 4d). Statistical significance was evaluated using a permutation method (see Methods). Very high decoding accuracy for temporal epoch identity across the whole session spanning both trial and intertrial periods was observed for all LEC animals during the BW12 experiment, indicating substantial temporal information was present in LEC population activity (88.0% mean accuracy, chance level 4.2%; Fig. 3a). The high decoding accuracy was not due to pooling cells across recording days, changes in behaviour or noise-driven variability of population activity states (Extended Data Figs. 4e, f, 5). To verify our observations further, we trained decoders using data from additional LEC animals that ran a simplified four-trial version of the BW12 experiment (BW4), and obtained similar results (93.1% mean accuracy, chance level 25%; Extended Data Fig. 4g, h). Thus, population activity in the LEC clearly defined a unique temporal context for every epoch of experience on the timescale of minutes.

Visualization of data from the CA3 and MEC revealed population activity patterns that were different from those observed in the LEC (Fig. 2). Quantification of temporal information across the entire session, including both trial and intertrial periods, using decoders trained on CA3 and MEC data revealed decoding accuracies above chance, but lower than that found in the LEC (23.0% and 34.4% mean accuracy for CA3 and MEC, respectively). To compare decoding accuracy properly across the three areas, we trained separate decoders for each area using populations with equal size to ensure that higher accuracy was not simply due to the data having higher dimensionality. The LEC still contained considerably more temporal information than either the CA3 or MEC, as decoding accuracy for temporal epoch was higher for the LEC by a sizeable margin (45.5%, 20.4% and 26.2% mean accuracies for LEC, CA3 and MEC, respectively; Fig. 3b, c). The difference in decoding accuracy was maintained across a range of population sizes, with the LEC requiring significantly fewer cells to reach high decoding accuracy (Fig. 3d, Extended Data Fig. 4i). In the simplified BW4 experiments, we also had data from the CA2 and CA1, allowing us to compare decoding accuracies for temporal epoch across the entorhinal cortex and all the CA subfields. We again found, in separate animals, that decoding accuracy was highest for the LEC (Extended Data Fig. 4h). In total, these observations point to the LEC as a possible source of temporal context information necessary for episodic memory formation in the hippocampus.

The robust encoding of time in the LEC at the population level could be due primarily to cells with ramp-like activity. To test whether cells

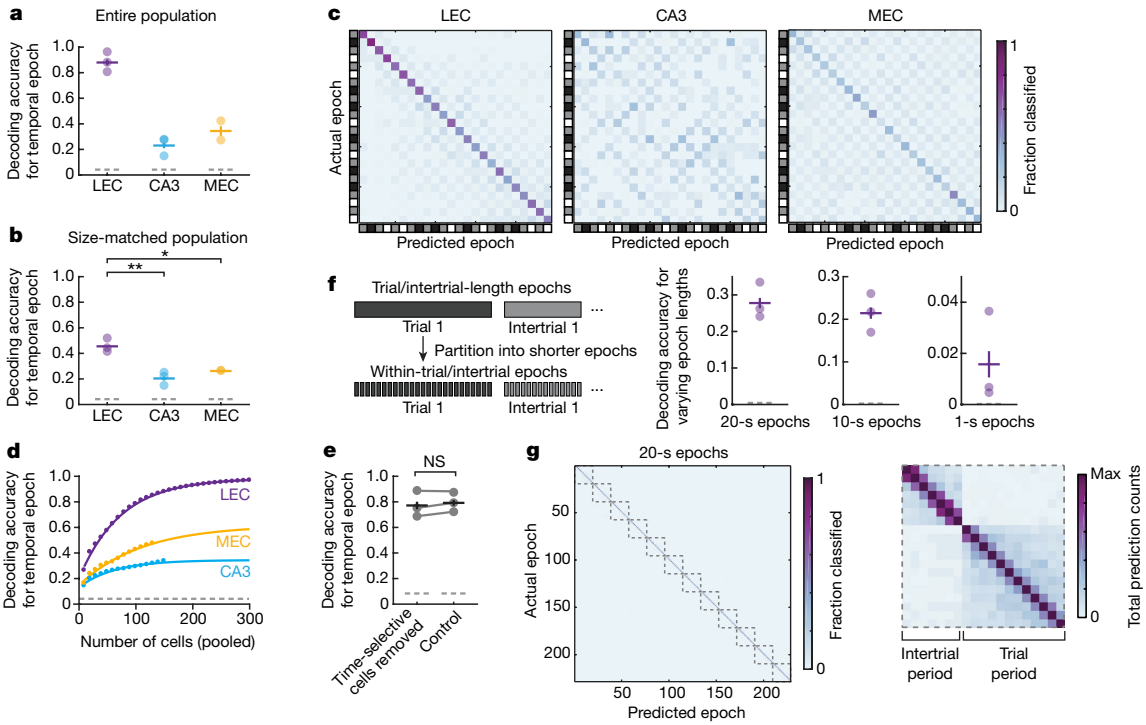


Fig. 3 | Temporal information within LEC population activity. **a**, Decoding accuracies for temporal epoch across the whole recording session for the LEC ($n = 3$ animals), CA3 ($n = 3$ animals), and MEC ($n = 2$ animals). **b**, As in **a**, but for size-matched populations. $^{*}P < 0.05$ (LEC versus MEC, $t(3) = 4.81$), $^{**}P < 0.01$ (LEC versus CA3, $t(4) = 5.77$), unpaired *t*-test with Bonferroni correction; matched population size = 28 cells. **c**, Confusion matrices from example animals, using size-matched populations. Each matrix shows the entire session, with each entry corresponding to a single epoch (trial 1, intertrial 1, trial 2 and so on). Epoch type is indicated along the left and bottom (grey for intertrial periods, black for black wall trial periods, white for white wall trial periods). **d**, Relationship between population size and decoding accuracy. Lines indicates exponential curve fit to data (shown as points, pooled across $n = 3$, 3 and 2 animals for LEC, CA3 and MEC, respectively). **e**, Decoding accuracy for temporal epoch using subpopulations with all time-selective cells removed, compared to randomly drawn size-matched controls. NS, not significant ($P = 0.42$; $t(2) = 1.00$), paired *t*-test; $n = 3$ animals with 62, 107 and 150 cells not selective for time. **f**, Decoding accuracies for different temporal epoch lengths for the LEC. Left, method for constructing shortened epochs. Right, decoding accuracies for different epoch lengths ($n = 3$ animals, matched population size = 90 cells). **g**, Confusion matrix for 20-s epochs from example animal. The matrix contains 228 epochs (each trial period truncated to 240 s and divided into 12 20-s epochs, each intertrial period of 140 s divided into 7 20-s epochs, 19 epochs per trial/intertrial pair, 12 total pairs across the session, giving 228 epochs). Left, confusion matrix for the entire session. Right, sum across trial/intertrial sections of the whole session matrix (outlined by dashed line in whole session matrix). **a**, **b**, **d**, **f**, Circles indicate individual animals, solid lines indicate mean decoding accuracies ± s.e.m., dashed lines indicate chance levels.

not classified as selective for time nonetheless also encoded temporal information, possibly in a nonlinear form[32], we trained decoders for temporal epoch using only cells that were not selective for time, and found that decoding accuracy remained very high. Comparing against size-matched populations with cells drawn randomly from the entire dataset, decoding accuracy for time using only non-time-selective cells was not significantly different (77.2% versus 79.2% mean accuracy for trial identity using a population with no time-selective cells, compared to a size-matched randomly drawn population; Fig. 3e).

If the temporal information observed within LEC population activity is actually used as a temporal code, an intriguing question is whether it is limited to representing just the macroscopic temporal context of experiences, on the timescale of minutes and longer, or whether it is also capable of representing the order of events within experiences, which may be on timescales shorter than minutes. We examined whether the LEC could encode temporal epochs of shorter length than entire trial or intertrial periods by dividing trial-period data into shorter epochs and training decoders to identify these shortened epochs (Fig. 3f). Decoding accuracies for 20-s, 10-s and 1-s epoch-lengths across the entire session were all significantly above chance (27.8%, 21.4% and 1.6% mean accuracy for 20-s, 10-s and 1-s epochs, respectively; chance levels 0.4%, 0.2% and 0.02%, respectively; Fig. 3f, g, Extended Data Fig. 6a–c and Supplementary Information). Although decoding accuracy decreased with shorter epoch lengths, decoding errors predominantly predicted adjacent epochs, as opposed to temporally distant epochs (Fig. 3g). Decreased accuracy may have been due to the limited number of cells that we recorded from. By pooling together all recorded cells from every animal to generate a population of 451 cells, the decoding accuracy for 20-s epochs was 86.0%, and the decoding accuracies for 10-s and 1-s epochs were 83.0% and 30.5%, respectively. Decoding accuracy for shortened epochs was not due to differences in population activity on the timescale of trial or intertrial periods, as decoding accuracy for shortened epochs within single trial and intertrial periods was still above chance (Extended Data Fig. 6d, e). Thus, temporal information in the LEC was flexible enough to support the encoding of events happening across a wide range of timescales.

In a separate set of experiments, we tested whether changing the content of the animal's experience affected temporal information in the LEC by introducing an object into the recording environment. We found a similar representation of time at both the single-cell and population level (Extended Data Fig. 7). Information about the current environmental context (B or W, Extended Data Fig. 4j–l) as well as information about the immediately preceding context was also present in the LEC (Supplementary Information). Overall, these results suggest a large component of LEC population activity codes for time across multiple scales, expressed both through cells explicitly coding time as well as cells having mixed selectivity for time that requires population decoding to extract.

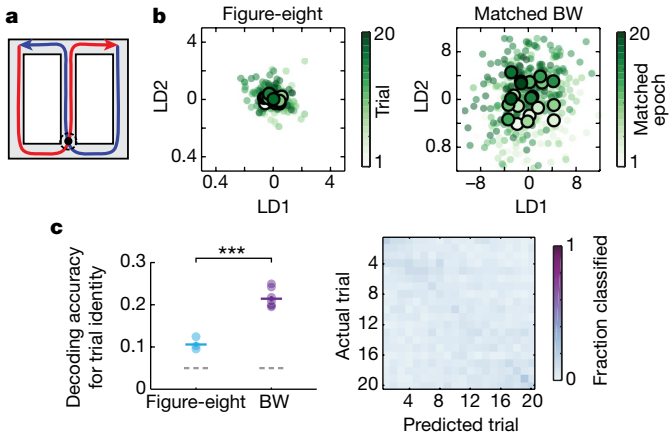


Fig. 4 | Temporal information arises inherently. a, Continuous alternation task, in which the animals alternated between left and right turns when they reached the top of the central stem for 40 total trials. Black circle indicates the base of the central stem. **b**, Left, 2D projection of LEC neural population responses during the figure-eight experiment from example animal. Right, 2D projection of LEC neural population responses during matched periods from BW experiments. **c**, Left, decoding accuracy for trial identity during the figure-eight experiment compared to decoding accuracy for temporal epoch using matched data from BW experiments. $***P < 0.0001$ ($t(8) = 8.35$), unpaired t-test; matched population size = 31 cells, $n = 3$ and 7 animals for figure-eight and BW, respectively. Circles indicate individual animals, solid lines indicate mean decoding accuracies ± s.e.m., dashed lines indicate chance levels. Right, confusion matrix for the figure-eight experiment from example animal.

Temporal information arises inherently

There are several possible mechanisms by which time may be incorporated into the population representation in the LEC. One possibility ('explicit mechanism') is that the LEC actively, in a clock-like manner, generates timestamps for representations of experience. Another possibility ('inherent mechanism') is that temporal information in the LEC arises simply because the animal's moment-to-moment experience constantly changes, and time can be extracted from this changing flow of experience by integrating the amount of change[9,12,33] (Extended Data Figs. 8, 9a). We sought to distinguish these two possibilities by constraining experience through the use of a more structured task in which the animal's behaviour was stereotyped. If the temporal information present within the LEC arose from an explicit clock-like process, we would expect to see no change in the amount of temporal information present when compared to results from the free-foraging BW12/BW4 experiments. By contrast, if the temporal information within the LEC arose inherently through its encoding of experience, we would expect to see a decrease in the amount of temporal information due to the repetitive nature of the task.

We examined the variability of temporal representation in the LEC in a separate experiment in which animals performed a learned continuous-alternation task (figure-eight task, Fig. 4a). Trials were aligned by the time point at which the animal entered the central stem, and trials consisted of activity spanning from three seconds before to three seconds after entering the central stem. Visualizing the data, we observed that the separation between trials was reduced in comparison to BW data (Fig. 4b). Decoding accuracy for trial identity across the entire session was above chance, but much lower than for matched BW data (10.6% versus 21.5% mean accuracy for figure-eight and BW respectively, chance level 5%; Fig. 4c, Extended Data Fig. 9b). In a separate

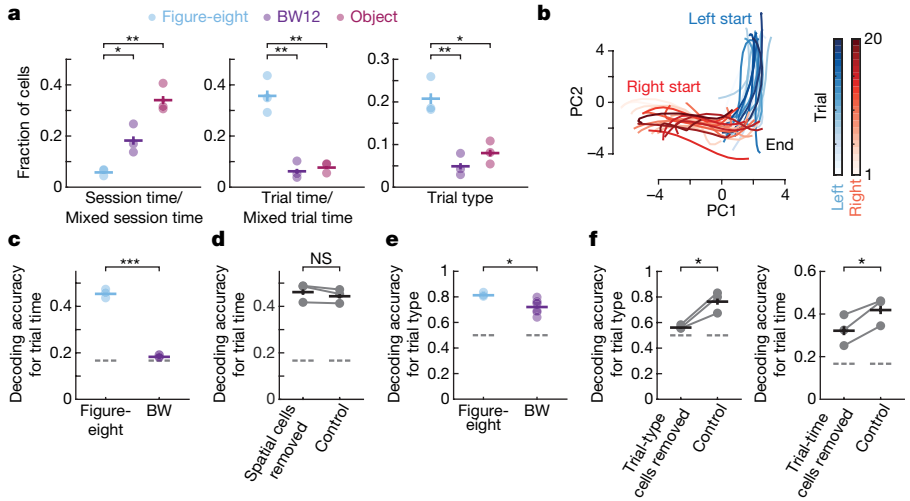


Fig. 5 | Temporal coding depends on behavioural context. a, Proportion of cells exhibiting selectivity for session time, trial time or trial type for the figure-eight (light blue), BW12 (purple), and object experiments (magenta). Session/mixed session time: $*P < 0.05$ (figure-eight versus BW12, $t(4) = 4.35$), $**P < 0.01$ (figure-eight versus object, $t(4) = 10.10$). Trial/mixed trial time: $**P < 0.01$ (figure-eight versus BW12, $t(4) = 6.70$; figure-eight versus object, $t(4) = 7.28$). Trial type: $**P < 0.01$ (figure-eight versus BW12, $t(4) = 5.56$), $*P < 0.05$ (figure-eight versus object, $t(4) = 4.45$), unpaired t-test on arcsine-transformed data with Bonferroni correction; $n = 3$ animals for all. **b**, 2D projection of neural trajectories for LEC data during the figure-eight experiment from example animal. Axes correspond to the first two principal components (PC1 and PC2; arbitrary units). **c**, Decoding accuracy for trial time compared to matched BW data. $***P < 10^{-10}$ ($t(8) = 38.98$), unpaired t-test; matched population size = 31 cells, $n = 3$ and 7 animals for figure-eight and BW, respectively. **d**, Decoding accuracy for trial time during the figure-eight experiment using subpopulations with all spatially selective cells removed, compared to randomly drawn size-matched controls. NS, not significant ($P = 0.16$, $t(2) = 2.17$), paired t-test; $n = 3$ animals with 38, 44 and 22 cells not selective for space. **e**, Decoding accuracy for trial type during the figure-eight experiment compared to decoding accuracy for wall colour using matched BW data. $*P < 0.05$ ($t(8) = 2.74$), unpaired t-test; matched population size = 31 cells, $n = 3$ and 7 animals for figure-eight and BW, respectively. **f**, Decoding accuracy for trial type (left) or trial time (right) during the figure-eight experiment, with cells selective for decoded variable removed, compared to randomly drawn size-matched controls. $*P < 0.05$ (trial type: $t(2) = 4.83$; trial time: $t(2) = 4.40$), paired t-test; $n = 3$ animals with 27, 29 and 8 cells not selective for trial time, and 29, 31 and 15 cells not selective for trial type). Circles indicate individual animals, solid lines indicate mean ± s.e.m. of described measurement, dashed lines indicate chance levels.

experiment in which animals ran repeated laps on a circular track, reduced decoding accuracy for trial identity during repetitive experience was also observed (Extended Data Fig. 9c–f). Overall, these results are consistent with temporal information in the LEC arising not in an explicit clock-like manner, but inherently from the dynamics underlying the representation of ongoing experience in the LEC.

Temporal coding varies with changing behaviour

The hippocampus represents learned time intervals[15,16,19,34], and this representation of time may depend on the retrieval of stable temporal contexts[35]. The reduction in across-session temporal information during learned behaviour may reflect a similar instatement of a stable temporal context in the LEC for the purpose of representing the relevant features of the task, which may include the progression of time within single trials. Analysis of single-cell responses using a GLM incorporating trial type (left/right turn), trial time (time within single trials), and session time showed that the ratio of cells encoding trial time versus session time was significantly altered, with more cells encoding trial time and fewer cells encoding session time in the figure-eight experiment (5.7%, 18.2% and 34.0% cells selective for session time or mixed session time (session time + trial type) for the figure-eight, BW12, and object experiments, respectively; 35.7%, 6.2% and 7.7% cells selective for trial or mixed trial time (trial time + trial type) for the figure-eight, BW12 and object experiments, respectively; Fig. 5a, Extended Data Fig. 10b–e). The proportion of cells selective for trial type also differed significantly (20.8%, 4.9% and 8.0% cells selective for context for the figure-eight, BW12 and object experiments, respectively; Fig. 5a).

Visualization of overall population activity using principal component analysis to plot 2D projections of neural trajectories through population activity space showed that neural trajectories were relatively constant across trials, suggesting that the LEC was in fact in a different, more stable mode of activity compared to that observed during free foraging (Fig. 5b, Extended Data Fig. 10f). Consistent with this observation, decoding accuracy for time relative to the start of each trial was significantly above chance and much higher than for matched BW data (45.3% versus 18.3% mean accuracy, chance level 16.6%; Fig. 5c). Although trial time was tightly correlated with position, the significant decoding of trial time did not seem to be due purely to LEC activity reflecting spatial location (10.0% cells selective for trial time exclusively with no influence of position, Fig. 5d). In addition to changes in the type of temporal information present in LEC population activity, the amount of task-related information also appeared to change, as the decoding accuracy for trial type was higher for the figure-eight data than the decoding accuracy for wall colour using matched BW data (81.3% versus 72.1% mean accuracy, chance level 50%; Fig. 5e). Finally, the degree to which information was distributed across the entire population was decreased compared to BW experiments, as decoding accuracy for both trial type and trial time were significantly reduced compared to size-matched randomly drawn controls when cells selective for trial type or trial time respectively were removed (56.1% versus 76.4% mean accuracy for trial type, 32.2% versus 41.9% mean accuracy for trial time; Fig. 5f). Overall, our results suggest that as animals engaged in a structured, learned task, the dynamics of LEC activity became considerably more stable compared to when animals were engaged in free behaviour.

Discussion

Being able to recall the temporal details of past experiences is a fundamental element of episodic memory. Our recordings demonstrate a unique temporal signal in the LEC that can encode time across multiple scales from seconds to hours and across different environmental contexts. Ordinarily this code for time marks the free-flowing progression of time, reflecting the structure of ongoing experience. However, when animals engage in a structured behavioural task in which experience is similar across repeated trials, time coding becomes relative, encoding time with respect to temporal landmarks. The adaptable nature of this code for time makes it particularly well-suited for defining the temporal component of episodic memory (episodic time) and differentiates it from previously described interval-timing mechanisms[36–38].

Our results support recent theoretical work demonstrating that encoding of time can arise inherently as a result of interactions between externally and internally driven states[31,33,39]. This form of time coding represents different points in time using different high-dimensional population states that can be easily differentiated by downstream readout neurons (Extended Data Fig. 8). In the LEC, such high-dimensional states may be generated by a combination of the recurrent local connectivity of the LEC[40] and the uniquely diverse set of inputs that the LEC receives[41]. Given the anatomical position of the LEC as a major gateway for information entering the hippocampus, it is possible that representations of time outside of the entorhinal/hippocampal circuit[42,43] are integrated in the LEC to form a single representation of time for episodic memories.

Within the entorhinal/hippocampal circuit, two representations of time have been identified in previous work: time cells, which fire at specific points in time as an animal performs a task[15–19], and the decorrelation of place cell activity across hours to days[20–25]. Continuously changing LEC activity may underlie both scales of temporal representation: drift of place cell activity may be governed by a constantly changing LEC input that is time-varying on the scale of minutes to hours, and sequential activity of time cells may be driven in part by LEC input that is time-varying relative to task events on the scale of seconds[31,44]. Thus, although an episodic memory may contain both a fine-grained representation of the sequence of events composing that memory as well as a coarser temporal context for the overall episode, both of these scales of temporal representation may originate from a single temporal signal within the LEC. This signal may then reach the hippocampus to become part of a unified what–when–where representation of experience, space and time[3], in which the representation of experience and time arising in the LEC is integrated with the representation of space arising in the MEC[45].

Online content

Any methods, additional references, Nature Research reporting summaries, source data, statements of data availability and associated accession codes are available at https://doi.org/10.1038/s41586-018-0459-6

Received: 25 September 2017; Accepted: 3 July 2018;
Published online 29 August 2018.

1. Tulving, E. *Elements of Episodic Memory* (Oxford Univ. Press, Oxford, 1985).
2. Clayton, N. S. & Dickinson, A. Episodic-like memory during cache recovery by scrub jays. *Nature* **395**, 272–274 (1998).
3. Eichenbaum, H. On the integration of space, time, and memory. *Neuron* **95**, 1007–1018 (2017).
4. Fortin, N. J., Agster, K. L. & Eichenbaum, H. B. Critical role of the hippocampus in memory for sequences of events. *Nat. Neurosci.* **5**, 458–462 (2002).
5. Kesner, R. P., Gilbert, P. E. & Barua, L. A. The role of the hippocampus in memory for the temporal order of a sequence of odors. *Behav. Neurosci.* **116**, 286–290 (2002).
6. Ekstrom, A. D. & Bookheimer, S. Y. Spatial and temporal episodic memory retrieval recruit dissociable functional networks in the human brain. *Learn. Mem.* **14**, 645–654 (2007).
7. Lehn, H. et al. A specific role of the human hippocampus in recall of temporal sequences. *J. Neurosci.* **29**, 3475–3484 (2009).
8. Naya, Y. & Suzuki, W. A. Integrating what and when across the primate medial temporal lobe. *Science* **333**, 773–776 (2011).
9. Ezzyat, Y. & Davachi, L. Similarity breeds proximity: pattern similarity within and across contexts is related to later mnemonic judgments of temporal proximity. *Neuron* **81**, 1179–1189 (2014).
10. Hsieh, L. T., Gruber, M. J., Jenkins, L. J. & Ranganath, C. Hippocampal activity patterns carry information about objects in temporal context. *Neuron* **81**, 1165–1178 (2014).
11. Dede, A. J. O., Frascino, J. C., Wixted, J. T. & Squire, L. R. Learning and remembering real-world events after medial temporal lobe damage. *Proc. Natl Acad. Sci. USA* **113**, 13480–13485 (2016).
12. Buzsáki, G. & Llinás, R. Space and time in the brain. *Science* **358**, 482–485 (2017).
13. Ferbinteanu, J., Kennedy, P. J. & Shapiro, M. L. Episodic memory—from brain to mind. *Hippocampus* **16**, 691–703 (2006).

14. Hassabis, D. & Maguire, E. A. Deconstructing episodic memory with construction. *Trends Cogn. Sci.* **11**, 299–306 (2007).
15. Pastalkova, E., Itskov, V., Amarasingham, A. & Buzsáki, G. Internally generated cell assembly sequences in the rat hippocampus. *Science* **321**, 1322–1327 (2008).
16. MacDonald, C. J., Lepage, K. Q., Eden, U. T. & Eichenbaum, H. Hippocampal "time cells" bridge the gap in memory for discontiguous events. *Neuron* **71**, 737–749 (2011).
17. Kraus, B. J. et al. During running in place, grid cells integrate elapsed time and distance run. *Neuron* **88**, 578–589 (2015).
18. Mau, W. et al. The same hippocampal CA1 population simultaneously codes temporal information over multiple timescales. *Curr. Biol.* **28**, 1499-1508.e4 (2018).
19. Eichenbaum, H. Time cells in the hippocampus: a new dimension for mapping memories. *Nat. Rev. Neurosci.* **15**, 732–744 (2014).
20. Cai, D. J. et al. A shared neural ensemble links distinct contextual memories encoded close in time. *Nature* **534**, 115–118 (2016).
21. Manns, J. R., Howard, M. W. & Eichenbaum, H. Gradual changes in hippocampal activity support remembering the order of events. *Neuron* **56**, 530–540 (2007).
22. Mankin, E. A. et al. Neuronal code for extended time in the hippocampus. *Proc. Natl Acad. Sci. USA* **109**, 19462–19467 (2012).
23. Ziv, Y. et al. Long-term dynamics of CA1 hippocampal place codes. *Nat. Neurosci.* **16**, 264–266 (2013).
24. Mankin, E. A., Diehl, G. W., Sparks, F. T., Leutgeb, S. & Leutgeb, J. K. Hippocampal CA2 activity patterns change over time to a larger extent than between spatial contexts. *Neuron* **85**, 190–201 (2015).
25. Rubin, A., Geva, N., Sheintuch, L. & Ziv, Y. Hippocampal ensemble dynamics timestamp events in long-term memory. *eLife* **4**, e12247 (2015).
26. Hargreaves, E. L., Rao, G., Lee, I. & Knierim, J. J. Major dissociation between medial and lateral entorhinal input to dorsal hippocampus. *Science* **308**, 1792–1794 (2005).
27. Deshmukh, S. S. & Knierim, J. J. Representation of non-spatial and spatial information in the lateral entorhinal cortex. *Front. Behav. Neurosci.* **5**, 69 (2011).
28. Lu, L. et al. Impaired hippocampal rate coding after lesions of the lateral entorhinal cortex. *Nat. Neurosci.* **16**, 1085–1093 (2013).
29. Keene, C. S. et al. Complementary functional organization of neuronal activity patterns in the perirhinal, lateral entorhinal, and medial entorhinal cortices. *J. Neurosci.* **36**, 3660–3675 (2016).
30. Pilkiw, M. et al. Phasic and tonic neuron ensemble codes for stimulus-environment conjunctions in the lateral entorhinal cortex. *eLife* **6**, e28611 (2017).
31. Howard, M. W. et al. A unified mathematical framework for coding time, space, and sequences in the hippocampal region. *J. Neurosci.* **34**, 4692–4707 (2014).
32. Rigotti, M. et al. The importance of mixed selectivity in complex cognitive tasks. *Nature* **497**, 585–590 (2013).
33. Buonomano, D. V. & Maass, W. State-dependent computations: spatiotemporal processing in cortical networks. *Nat. Rev. Neurosci.* **10**, 113–125 (2009).
34. McEchron, M. D., Tseng, W. & Disterhoft, J. F. Single neurons in CA1 hippocampus encode trace interval duration during trace heart rate (fear) conditioning in rabbit. *J. Neurosci.* **23**, 1535–1547 (2003).
35. Folkerts, S., Rutishauser, U. & Howard, M. W. Human episodic memory retrieval is accompanied by a neural contiguity effect. *J. Neurosci.* **38**, 4200–4211 (2018).
36. Finnerty, G. T., Shadlen, M. N., Jazayeri, M., Nobre, A. C. & Buonomano, D. V. Time in cortical circuits. *J. Neurosci.* **35**, 13912–13916 (2015).
37. Mauk, M. D. & Buonomano, D. V. The neural basis of temporal processing. *Annu. Rev. Neurosci.* **27**, 307–340 (2004).
38. Merchant, H., Harrington, D. L. & Meck, W. H. Neural basis of the perception and estimation of time. *Annu. Rev. Neurosci.* **36**, 313–336 (2013).
39. Howard, M. W. & Kahana, M. J. A distributed representation of temporal context. *J. Math. Psychol.* **46**, 269–299 (2002).
40. Dolorfo, C. L. & Amaral, D. G. Entorhinal cortex of the rat: organization of intrinsic connections. *J. Comp. Neurol.* **398**, 49–82 (1998).
41. Bota, M., Sporns, O. & Swanson, L. W. Architecture of the cerebral cortical association connectome underlying cognition. *Proc. Natl Acad. Sci. USA* **112**, E2093–E2101 (2015).
42. Jin, D. Z. Z., Fujii, N. & Graybiel, A. M. Neural representation of time in cortico-basal ganglia circuits. *Proc. Natl Acad. Sci. USA* **106**, 19156–19161 (2009).
43. Hyman, J. M., Ma, L., Balaguer-Ballester, E., Durstewitz, D. & Seamans, J. K. Contextual encoding by ensembles of medial prefrontal cortex neurons. *Proc. Natl Acad. Sci. USA* **109**, 5086–5091 (2012).
44. Igarashi, K. M., Lu, L., Colgin, L. L., Moser, M. B. & Moser, E. I. Coordination of entorhinal-hippocampal ensemble activity during associative learning. *Nature* **510**, 143–147 (2014).
45. Rowland, D. C., Roudi, Y., Moser, M. B. & Moser, E. I. Ten years of grid cells. *Annu. Rev. Neurosci.* **39**, 19–40 (2016).

Acknowledgements We thank A. M. Amundsgård, K. Haugen, K. Jenssen, E. Kråkvik, and H. Waade for technical assistance and M. P. Witter for help with determining recording locations in the LEC. The work was supported by a fellowship from the Helen Hay Whitney Foundation (to A.T.), an Advanced Investigator Grant from the European Research Council (GRIDCODE', grant number 338865), the Centre of Excellence scheme and the National Infrastructure Scheme of the Research Council of Norway (Centre for Neural Computation, grant number 223262; NORBRAIN1, grant number 197467), the Louis Jeantet Prize, the Körber Prize, and the Kavli Foundation.

Reviewer information *Nature* thanks M. Shapiro and the other anonymous reviewer(s) for their contribution to the peer review of this work.

Author contributions A.T., M.-B.M. and E.I.M. designed experiments and the analytic approach; A.T. developed and performed analyses; A.T., J.S. and L.L. collected data; C.W. and J.J.K. contributed circular track data. A.T. and E.I.M. wrote the paper with input from all authors. A.T. wrote the first draft and had the main role in developing the paper.

Competing interests The authors declare no competing interests.

Additional information
Extended data is available for this paper at https://doi.org/10.1038/s41586-018-0459-6.
Supplementary information is available for this paper at https://doi.org/10.1038/s41586-018-0459-6.
Reprints and permissions information is available at http://www.nature.com/reprints.
Correspondence and requests for materials should be addressed to A.T. or E.I.M.
Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

METHODS

Subjects. Experiments were carried out using twenty-one male Long Evans rats at NTNU and two male Long-Evans rats at Johns Hopkins University. Animals were housed individually in Plexiglas cages. Ten rats had tetrodes in the LEC, two rats had tetrodes in the LEC and CA3, two rats had tetrodes in the MEC, and nine rats had tetrodes in the CA3, CA2 or CA1. Data from seven of the LEC animals and the nine animals with tetrodes solely in the hippocampus have been published previously, but examined activity of individual cells rather than the overall population[46,47]. Animals were maintained on a 12-h light/12-h dark schedule, and kept at 85–90% of free-feeding body weight. Experiments were performed in accordance with the Norwegian Animal Welfare Act and the European Convention for the Protection of Vertebrate Animals used for Experimental and Other Scientific Purposes, or the Institutional Animal Care and Use Committee at Johns Hopkins University. The study contained no randomization to experimental treatments and no blinding. Sample size (the number of animals) was set based on conventions in the field.

Surgery and electrode preparation. Rats were anaesthetized with isoflurane (air flow: 1.0 l min^{-1}, 0.5–3% isoflurane, adjusted according to physiological monitoring). For LEC recordings, either one microdrive with four tetrodes (10 rats) or one hyperdrive with 18 tetrodes was targeted to the LEC (2 rats). Coordinates for microdrive LEC animals were: anterior-posterior (AP): 0–0.2 mm anterior to lambda, medial-lateral (ML): 5.3–5.5 mm lateral to midline, dorsal-ventral (DV): 4.3–5.0 mm below dura, 4–8° angle in the coronal plane, with electrode tips pointing away from midline. Coordinates for hyperdrive LEC animals were: AP: 7.5–7.6 mm posterior to bregma, ML: 3.0 mm lateral to midline, DV: at brain surface, 25° angle in the coronal plane, with electrode tips pointing away from midline. For MEC recordings, one microdrive was targeted to the MEC with coordinates: AP: 0.2–0.4 mm anterior to the transverse sinus, ML: 4.6 mm lateral to midline, DV: 1.8 mm below dura, 20° angle in the sagittal plane, with electrode tips pointing towards bregma. For hippocampal recordings, either one microdrive was targeted to the CA3, with coordinates: AP: 3.8 mm posterior to bregma, 3.0 mm lateral to midline, DV: 1.5 mm below dura (3 rats), or one hyperdrive was placed over hippocampus with coordinates: AP: 2.5–4.3 mm posterior to bregma, ML: 2.2–3.8 mm lateral to midline (9 rats). Drives were fixed to the skull using jeweller's screws and dental cement. One screw served as a ground for each drive. Tetrodes were made from four twisted 17 μm polyimide-coated platinum-iridium (90–10%) wires (California Fine Wire). Electrode tips were plated with platinum to reduce electrode impedances to between 150–300 kΩ at 1 kHz.

Recording procedures. Microdrives were connected to a multi-channel unity gain headstage, which was connected via a cable to a Neuralynx recording system for BW12 and circular-track experiments, or an Axona recording system for the BW4, object and figure-eight experiments (Axona Ltd). Unit activity was amplified by a factor of 1,000–10,000 and band-pass filtered from 600 to 6,000 Hz for Neuralynx recordings or from 800 to 6,700 Hz for Axona recordings. Spike waveforms above a set threshold were time-stamped and digitized at 32 kHz for 1 ms for all recordings. Tetrodes were lowered in 50-μm steps while the rat rested on a towel in a flower pot on a pedestal. Turning stopped when well-separated units appeared. Data collection started when signal amplitudes exceeded approximately five times the noise level (root mean square 20–50 μV) and units were stable for >3 h. Animal position was tracked at 25–30 Hz for Neuralynx recordings or 50 Hz for Axona recordings using an overhead video camera and either two (for microdrive animals) or multiple (for hyperdrive animals) LEDs attached to the headstage.

Behavioural procedures. For BW12, BW4 and object experiments, animals were trained to collect randomly scattered chocolate cereal crumbs in a square box. For BW12 experiments, animals ran in an 80 × 80 × 50 cm box with interchangeable walls, a single cue card along one wall, and curtains on two sides of the box. For BW4 experiments, the box was 100 × 100 cm. For object experiments, the box was 100 × 100 cm, the walls were black with a single cue card, and no curtains were present, such that there were many distal cues. Once animals were able to achieve good coverage of the recording environment, they began running standard sessions of 12 trials for BW12 experiments or 4 trials for BW4 experiments, during which the wall colour was changed across trials. For object experiments, animals ran 3 trials, during which an object consisting of a 6 × 6 × 37 cm tower made of Lego was placed in a fixed location during the second trial[47].

For circular track experiments, animals were trained to run back and forth between food wells for food pellets (BioServ) on a circular track (diameter 97 cm, width 10 cm). The food wells were separated by a 15 cm tall black barrier with 0.4 cm sidewalls. For the figure-eight experiments, animals were trained to run in a figure-eight pattern on a square figure-eight maze made of plexiglass with vinyl flooring which had runways 15 cm wide with 2 cm high side walls, a central stem that was 150 cm long, and total dimensions of 150 × 150 cm. Rewards were placed in dishes at the end of each goal arm. The maze was elevated 50 cm above the ground, surrounded by black curtains with a cue card 100 × 25 cm on the left side of the maze. Animals were trained in three stages[48]. In the first stage, wood-block barriers were placed at the bottom and top of the central stem so that rats could only run a fixed path. The position of the barriers was alternated on each trial once the rat reached the reward zone, and animals remained at this stage until they ran >20 trials within a 20-min session. In the second stage, the barrier at the top of the central stem was phased out so that animals could enter either reward arm. Only correct alternating choices were rewarded, and animals were blocked from backtracking to explore the stem or other reward arm once a choice was made. In the third and final stage, the barrier at the bottom of the central stem was phased out, and animals had to continue to run in a figure-eight pattern, with backtracking blocked by a barrier. After reaching the third stage, animals ran 20-min sessions, averaging ≥50 trials per session with >95% correct choices.

Spike sorting and cell classification. Spike sorting was performed offline using graphical cluster-cutting software to examine two-dimensional projections of the multidimensional parameter space (Tint (N. Burgess) for Axona data; MClust (A. D. Redish) or custom-written spike-sorting software for Neuralynx data). Spike clusters were compared across successive days to ensure that the same cell was not counted twice.

Cell stability across BW12 experimental sessions was determined by checking for stationarity of spike waveforms. For each cell, the Euclidean distance between spike waveforms across the recording session and the average waveform of the first ten spikes was measured. For each spike, the four spike waveforms were compressed into a single point in a 128-dimensional space (4 waveforms, each with 32 values), and Euclidean distance was measured in this 128-dimensional space. Subsequently, the stationarity of this distance was determined using three tests: augmented Dickey–Fuller test, the Kwiatkowski–Phillips–Schmidt–Shin test, and the Ljung–Box test, with stable cells being those which were categorized as stationary across all three tests. Following this stability test, stability was also assessed by calculating the Pearson correlation between the measured Euclidean distances across the recording session and firing rate across the recording session. Cells were excluded from further analysis if their correlation values exceeded chance levels. Chance levels were estimated individually for each cell by temporally shuffling the firing rate for a given cell and calculating the correlation between the waveform distance of the cell and the shuffled firing rates, repeating this process 1,000 times, and taking the 95th percentile as the threshold. Firing rate was estimated for 500-ms bins, with no smoothing or additional pre-processing.

GLM fitting. A Poisson GLM was fit to each cell individually using the MATLAB stepwiseglm function. The variables used to fit the model for BW12 data were wall colour, position, trial time, and session time. A single predictor was used for each variable. Session time was the total elapsed time, and took into account intertrial periods. Adjusted R^2 was used as the criterion for adding or removing terms (0.01 for adding, 0.005 for removing). Firing rate was estimated for 500-ms bins, with no smoothing or additional preprocessing. The same process was used for the object and figure-eight data, excluding position as a variable.

Estimating time constants. For cells that were classified as selective for trial or session time, a single-term exponential model was fit to the firing rate of each cell, $f = ae^{bx}$, in which f is the firing rate of the cell, a and b are constant coefficients, and b is the time constant of the cell. For cells selective for session time, the firing rate across all trial periods was used. For cells selective for trial time, the average firing rate across trials was used.

Dimension-reduced visualization of data. 2D projections of population activity states were constructed by reducing the dimensionality of raw neural data using principal component analysis (PCA), followed by applying linear discriminant analysis to the dimension-reduced data and taking the top two linear discriminants. Dimension reduction for visualization was carried out separately for 'whole session' and 'trial periods only' plots. For neural trajectories, PCA was applied to spiketrains that were first smoothed using a Gaussian kernel 500 ms wide. The top two principal components were taken for 2D projections. In both cases, neural data were from single animals but pooled across recording days. Bin size was 500 ms for both projections of population activity states and neural trajectories.

Decoding analysis. Decoding was done with linear support vector machine classifiers implemented using the LIBLINEAR package[49], specifically using L2-regularized L2-loss SVC, tenfold cross-validation, and cost parameter $C = 1$. Data consisted of raw spike counts from individual animals, pooled across recording days, and binned in 10-s bins, except for results varying epoch length, circular track, and figure-eight experiments. For shortened epochs, bin size was 1 s for 20-s-long epochs, 500 ms for 10-s-long epochs, and 50 ms for 1-s-long epochs. For the circular track and figure-eight results, the bin size was 500 ms. Multiclass decoding was done using the one-versus-all method. The cost parameter was varied between 10^{-3} and 10^{3} with no significant effect. Tenfold cross-validation was implemented by splitting the data into 10 subsamples, each randomly drawn from across the entire session. Decoders were then trained on 9 of the subsamples, and tested on the remaining subsample, with this process repeated using all 10 subsamples were used as test data once (Extended Data Fig. 4d). Bin-size was sufficiently shorter than epoch length to ensure that training data covered all

temporal epochs. Decoding accuracy was taken as the average accuracy across all 10 trained decoders. Cross-validation was repeated 1,000 times with overall decoding accuracy taken as the mean across the 1,000 repetitions, except for when population subsampling occurred, in which case 1,000 different subsamples were taken, each with 10 repetitions of cross-validation. For comparing decoding accuracies across areas, size-matched populations were used. Size-matched populations were generated by subsampling the total population of cells for a given animal without replacement. For each animal, subsampling was repeated 1000 times, and the mean was taken as the decoding accuracy for that animal. For decoding analyses in which GLM-classified cells were removed, the size-matched populations for comparison were randomly drawn from the full population of cells recorded for each individual animal without replacement.

For the figure-eight and circular track experiments, matched BW data consisted of 6-s-long epochs with 500 ms bins and an intertrial interval of 22 s. Because inclusion of a matched intertrial interval caused matched data to exceed the length of a single BW trial, matched data spanned across 3 trials of BW data. Wall colour in matched BW data corresponded to trial type for the figure-eight data (black walls/left trial, white walls/right trial), and running direction for circular track data (black walls/clockwise, white walls/anticlockwise). In total, 20 epochs were used from the BW data to match the 20 trials from the figure-eight experiment, with population size set to 31 cells. For circular track data, 15 epochs were used with population size set to 47 cells. Figure-eight and circular track data were also compared against matched BW data that did not account for the intertrial period. In this case, matched data were generated by taking activity from the first trial only of each session, with matched bin size and population size (20 epochs, each 6-s long, made up of 12 500-ms bins, in total spanning the first 120 s of the first trial, with population size set to 31 cells for the figure-eight data; 15 epochs, each 6-s long, made up of 12 500-ms bins, in total spanning the first 90 s of the first trial, with population size set to 47 cells for circular track data).

Determining statistical significance for decoding accuracies. Statistical significance for decoding accuracy was determined by comparing mean decoding accuracy from the original data against mean decoding accuracy from temporally shuffled data. Shuffled comparisons were generated by first temporally shuffling population activity such that the shuffled order was the same for each cell within the population. Decoding accuracy was then determined using the original labels for time bins and tenfold cross-validation. This process was repeated 1,000 times to generate shuffled comparisons for full-population decoding accuracies, and 1,000 times for each subsampled population of cells to generate shuffled comparisons for size-matched decoding accuracies. All decoding accuracies were significantly above chance, with the least significant decoding accuracy having $P < 10^{-5}$.

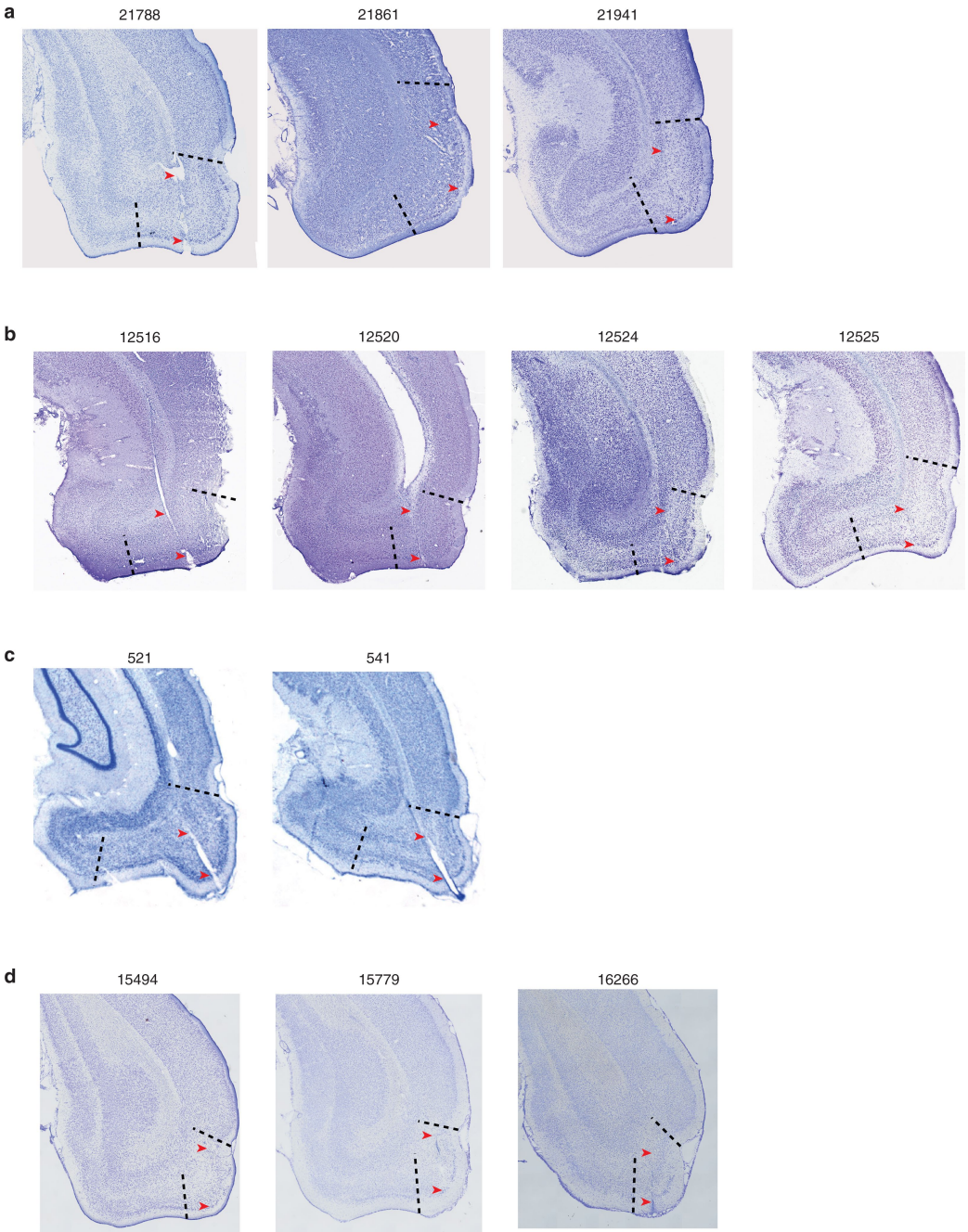
Estimates of distance between high-dimensional population states. For within-trial or intertrial period distances, population states were defined by the firing rates of cells from individual animals, pooled across recording days, for 10-s time bins. Firing rates were not smoothed. Distances were measured for each of the 12 trials or intertrials, and the average was taken as the final result for each animal. For across-trial or intertrial period distances, population states were defined in the same way, but then the average population state for each trial or intertrial period was calculated. Because the data were high-dimensional, Manhattan distance was used instead of Euclidean distance[50]. To account for differences in dimensionality across animals and facilitate comparison, distance measures were z-scored.

Classifying spatial cell types and measuring spatial tuning. Spatial rate maps were generated by binning activity into 3×3 cm spatial bins, calculating firing rates for each spatial bin, and then applying a two-dimensional Gaussian kernel with standard deviation of 7 cm in both directions. Only spikes recorded during running speeds above 2.5 cm s^{-1} were used. Place cells were classified by comparing spatial information scores against a shuffled distribution[51]. Grid cells were classified by comparing autocorrelogram-based gridness scores against a shuffled distribution[51]. Speed cells were classified by comparing Pearson correlation values against a shuffled distribution[52]. Velocity, acceleration and head direction were calculated using methods published previously[52]. Spatial selectivity for LEC cells was determined by measuring spatial information[53] for each cell and comparing against a distribution generated by calculating spatial information for data shuffled temporally using wrap-around shuffling.

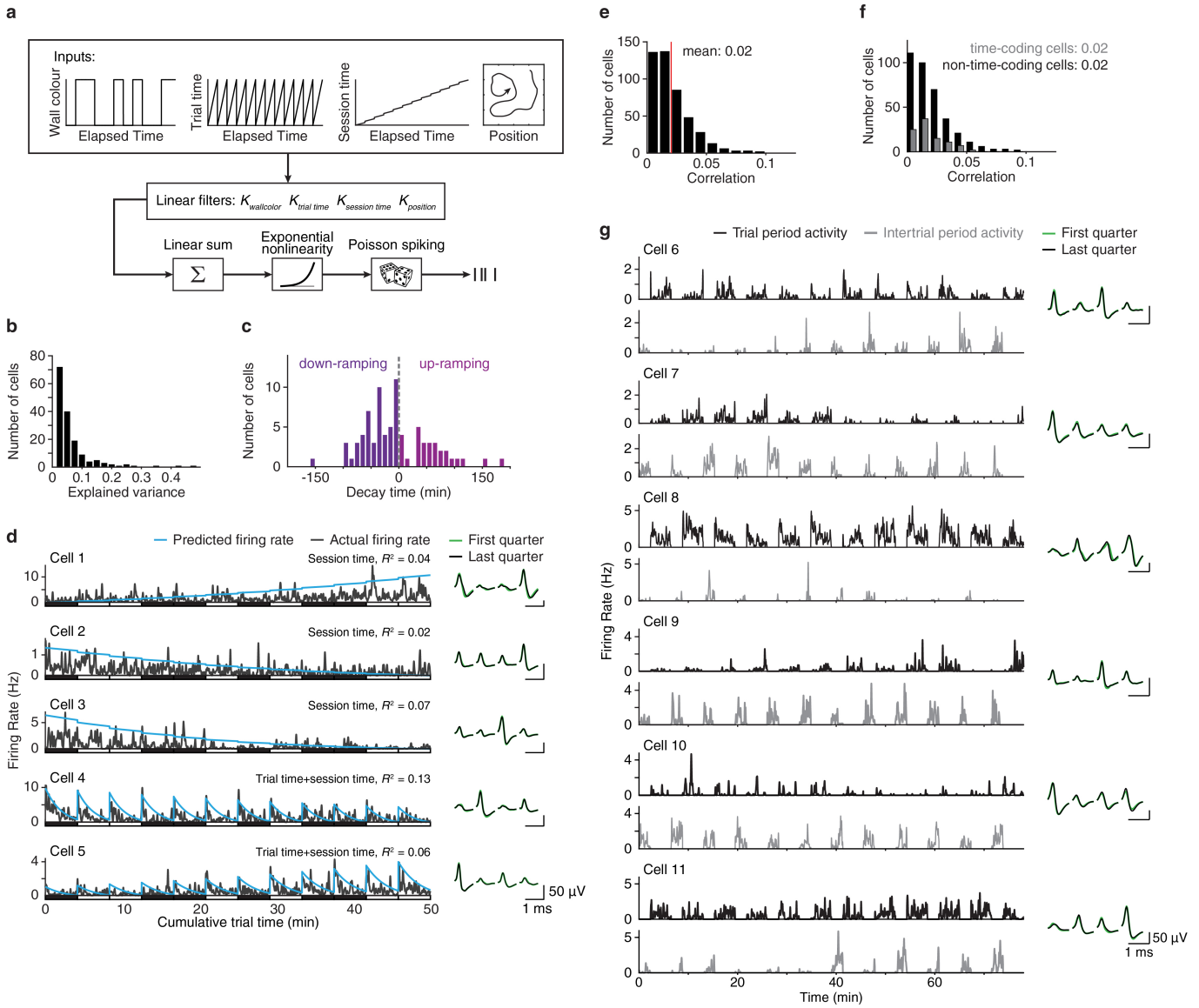
Reporting summary. Further information on experimental design is available in the Nature Research Reporting Summary linked to this paper.

Code and data availability. Custom code used and datasets generated and/or analysed during the current study are available from the corresponding author upon request.

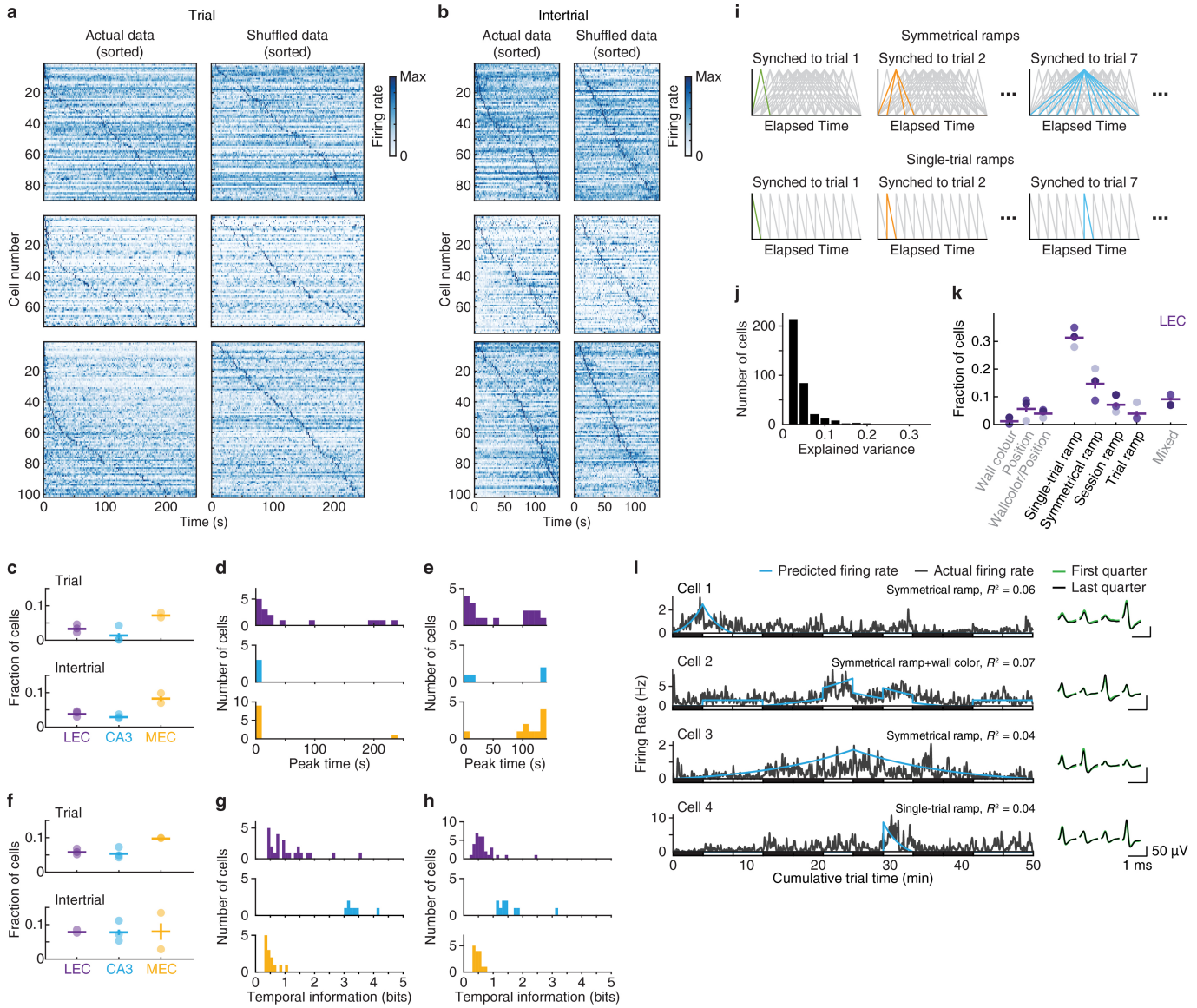
46. Lu, L., Igarashi, K. M., Witter, M. P., Moser, E. I. & Moser, M. B. Topography of place maps along the CA3-to-CA2 axis of the hippocampus. *Neuron* **87**, 1078–1092 (2015).
47. Tsao, A., Moser, M. B. & Moser, E. I. Traces of experience in the lateral entorhinal cortex. *Curr. Biol.* **23**, 399–405 (2013).
48. Wood, E. R., Dudchenko, P. A., Robitsek, R. J. & Eichenbaum, H. Hippocampal neurons encode information about different types of memory episodes occurring in the same location. *Neuron* **27**, 623–633 (2000).
49. Fan, R. E., Chang, K. W., Hsieh, C. J., Wang, X. R. & Lin, C. J. Liblinear: A library for large linear classification. *J. Mach. Learn. Res.* **9**, 1871–1874 (2008).
50. Aggarwal, C. C., Hinneburg, A. & Keim, D. A. On the surprising behavior of distance metrics in high dimensional space. *Lect. Notes Comput. Sci.* **1973**, 420–434 (2001).
51. Langston, R. F. et al. Development of the spatial representation system in the rat. *Science* **328**, 1576–1580 (2010).
52. Kropff, E., Carmichael, J. E., Moser, M. B. & Moser, E. I. Speed cells in the medial entorhinal cortex. *Nature* **523**, 419–424 (2015).
53. Skaggs, W. E., McNaughton, B. L. & Gothard, K. M. in *Advances in Neural Information Processing Systems 5 (NIPS 1992)* 1030–1037 (Morgan Kaufmann Publishers Inc., 1993).



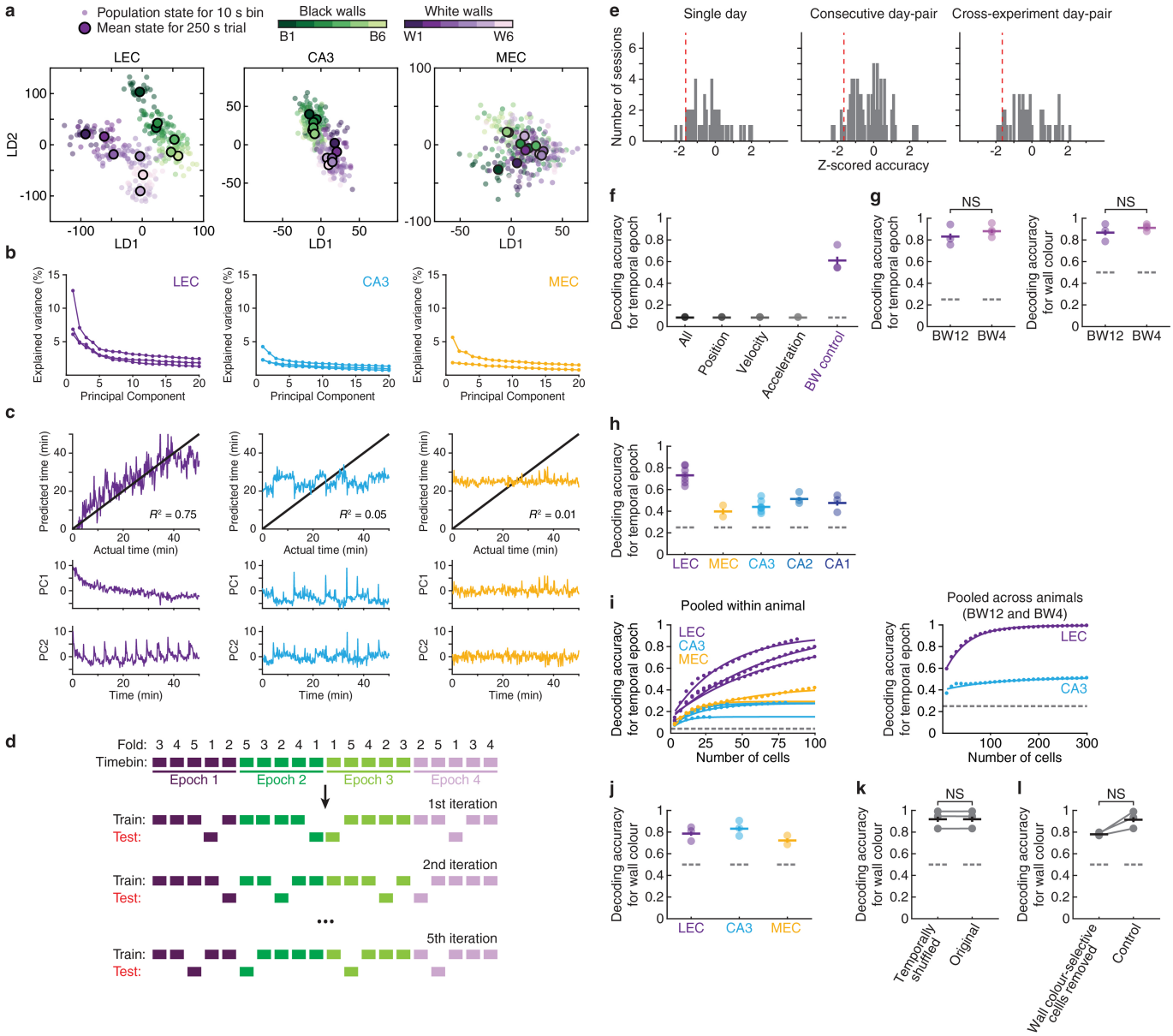
Extended Data Fig. 1 | Histology for LEC animals. a–d, Nissl-stained coronal sections showing recording locations for LEC animals used in the BW12 experiments (**a**), BW4 and figure-eight experiments (**b**), circular track experiments (**c**), and object experiments (**d**). Red arrowheads indicate range for tetrode locations. Activity was recorded from neurons in all layers of the lateral half of the LEC. Dashed lines indicate approximate anatomical borders of the LEC.



Extended Data Fig. 2 | Single-cell responses in the LEC related to time. **a**, Schematic of the GLM used for determining cell selectivity. Learned weights for the relevant predictors, as determined by the stepwise selection process, were put through an exponential nonlinearity that returned the mean rate of a Poisson process from which spikes were drawn (for LEC, 14.0% selective for session time, 3.8% selective for trial time, 2.7% selective for a mixture of trial and session time, percentages averaged across all animals). **b**, Explained variance for all LEC cells fit by the GLM for BW12 experiment ($n = 186$ cells). Average explained variance was 0.05. **c**, Distribution of time constants for trial-time or session-time selective cells ($n = 80$ cells). Time constants were estimated for each cell classified as trial-time or session-time selective by fitting a single-term exponential. **d**, Left, five additional example cells exhibiting ramping activity across the session. The firing rate of each cell is shown in grey, with the model-predicted firing rate in blue. The R^2 value is shown in for each cell. Right, average waveforms across four recording channels for the first (green line) and last (black line) quarter of the session. **e**, Pearson correlation between waveform distances (Euclidean distance between pairs of consecutive spike waveforms) and firing rate for all LEC cells used in the BW12 experiment ($n = 451$ cells). **f**, Comparison of the correlation between waveform distance and firing rate, as in **e**, for time-selective LEC cells ($n = 92$ cells, grey) and all other LEC cells used in the BW12 experiment ($n = 359$ cells, black). **g**, Six example cells demonstrating that notable fluctuations in firing rate can occur during stable recordings. For each cell, the top left row shows activity during trial periods, bottom left row shows activity during intertrial periods, and right panel shows average waveforms, as in **d**.

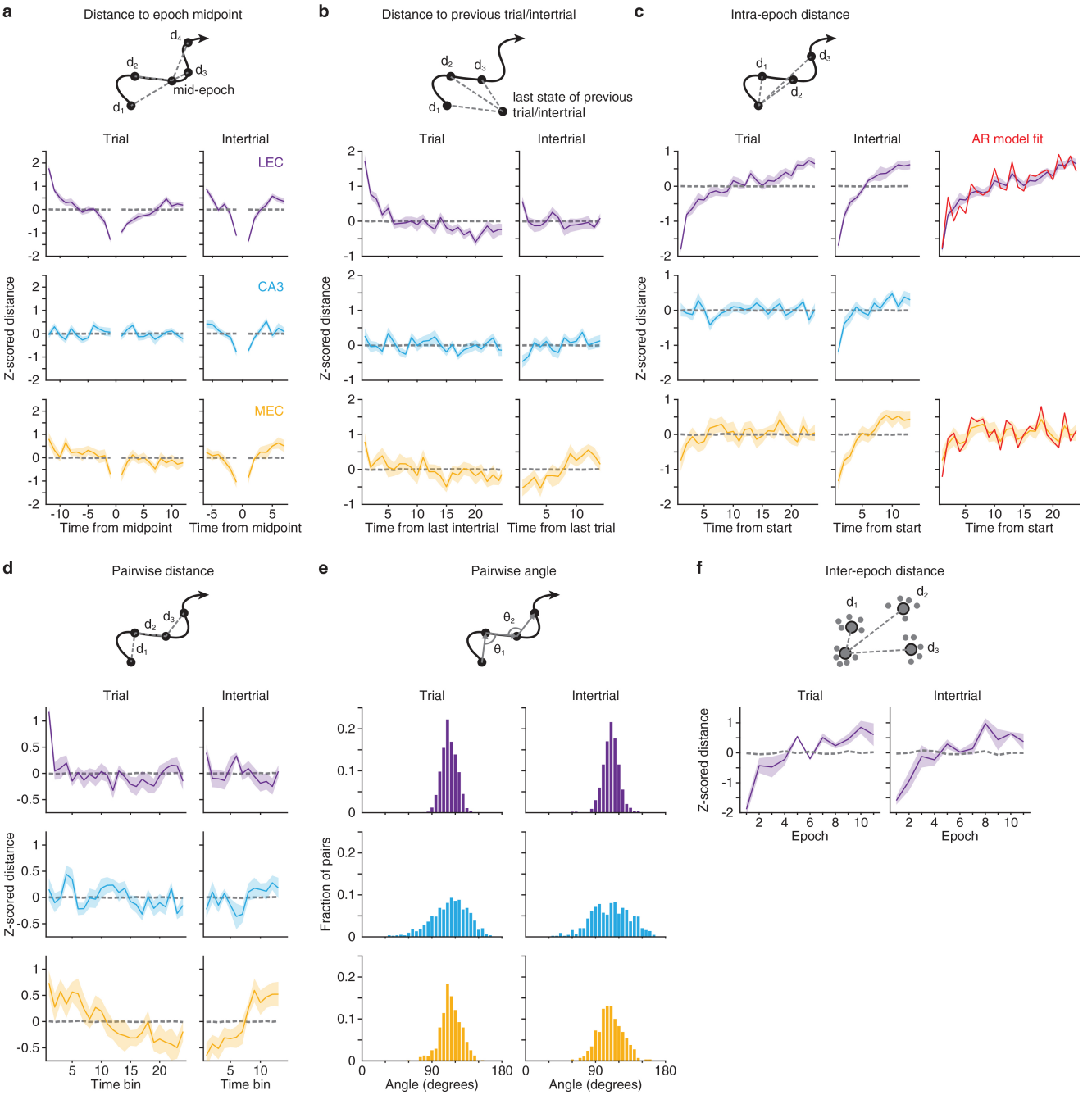


Extended Data Fig. 3 | Time-cell-like activity. a–h, Temporal specificity examined as a function of peak firing times of individual cells. **a**, **b**, Mean activity from example animals for the LEC (top matrix), CA3 (middle matrix) and MEC (bottom matrix) during trial (**a**) and intertrial (**b**) periods. For each matrix, rows show mean firing rates for individual cells ordered by the time of peak firing rate. Actual data shown in left column, shuffled data shown in right column. Cell identities were not maintained across actual and shuffled data. **c**, Fraction of cells with significant temporally specific activity for trial (top) and intertrial (bottom) periods ($n = 3$, 3 and 2 animals for LEC, CA3 and MEC, respectively). **d**, Time of peak activity for cells with significant temporally specific activity during trial periods. **e**, As in **d**, but for intertrial periods. **f–h**, Temporal specificity examined by calculating temporal information. **f**, Fraction of cells with significant temporal information for trial (top) and intertrial (bottom) periods ($n = 3$, 3 and 2 animals for LEC, CA3 and MEC, respectively). **g**, Distribution of significant temporal information scores for trial periods. **h**, As in **g**, but for intertrial periods. **i**, Predictors added for expanded GLM: symmetrical ramps and single-trial ramps. Coloured lines highlight the predictors for the respective trial indicated on top, grey lines are the rest of the available predictors. **j**, Explained variance for all LEC cells fit by expanded GLM for the BW12 experiment ($n = 350$ cells). Average explained variance was 0.03. **k**, Distribution of selectivity for time, wall colour and position for single LEC cells, determined using the expanded GLM ($n = 3$ animals). Shade indicates the same individual across the different variables. **l**, Left, examples of expanded GLM fit results for four cells with selectivity for different features. The firing rate of each cell is shown in grey, with the model-predicted firing rate in blue. The R^2 value is shown for each cell. Right, average waveforms across four recording channels for the first quarter (green line) and last quarter (black line) quarter of the session. Circles indicate individual animals, solid lines indicate mean fraction of cells ± s.e.m.

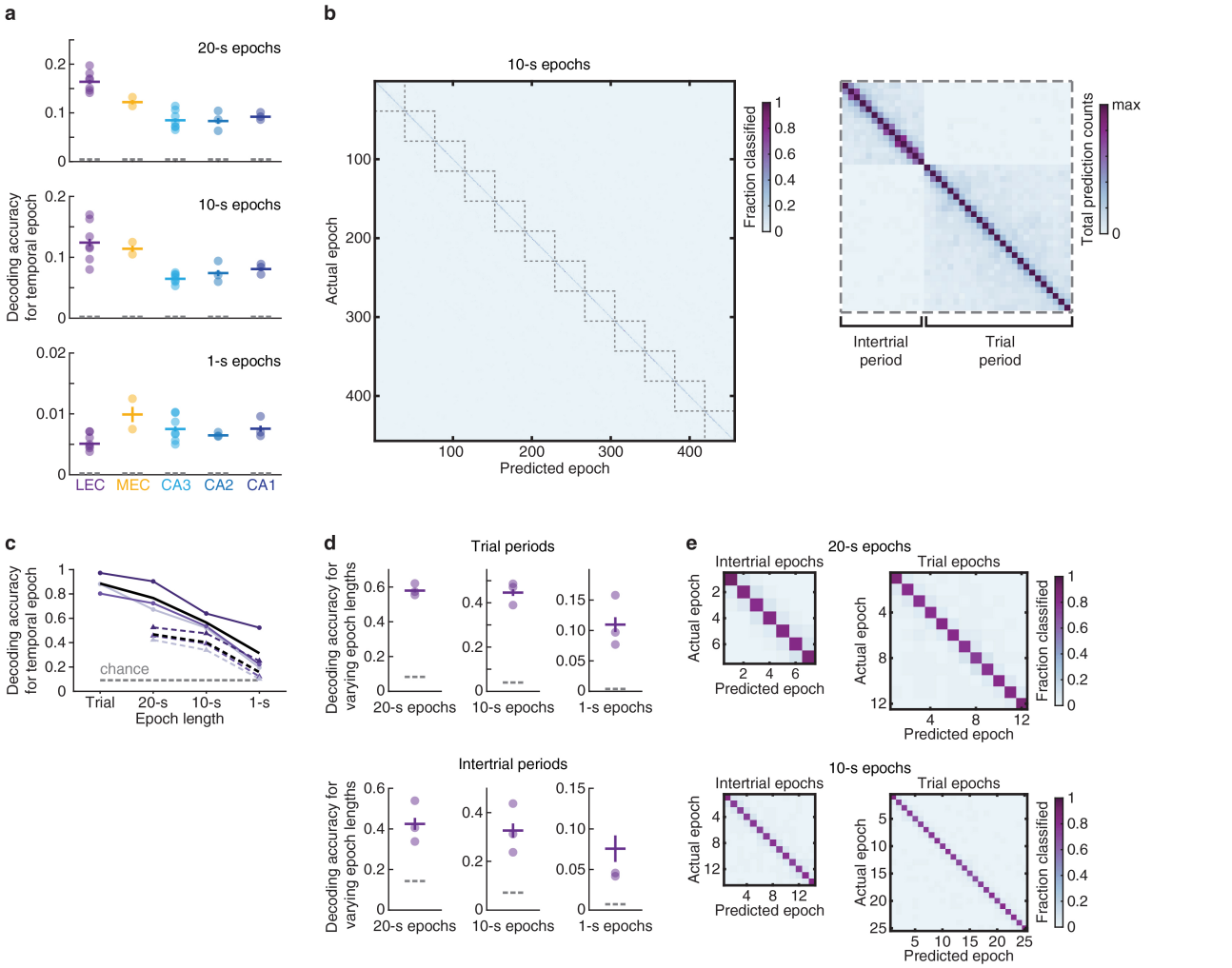


Extended Data Fig. 4 | See next page for caption.

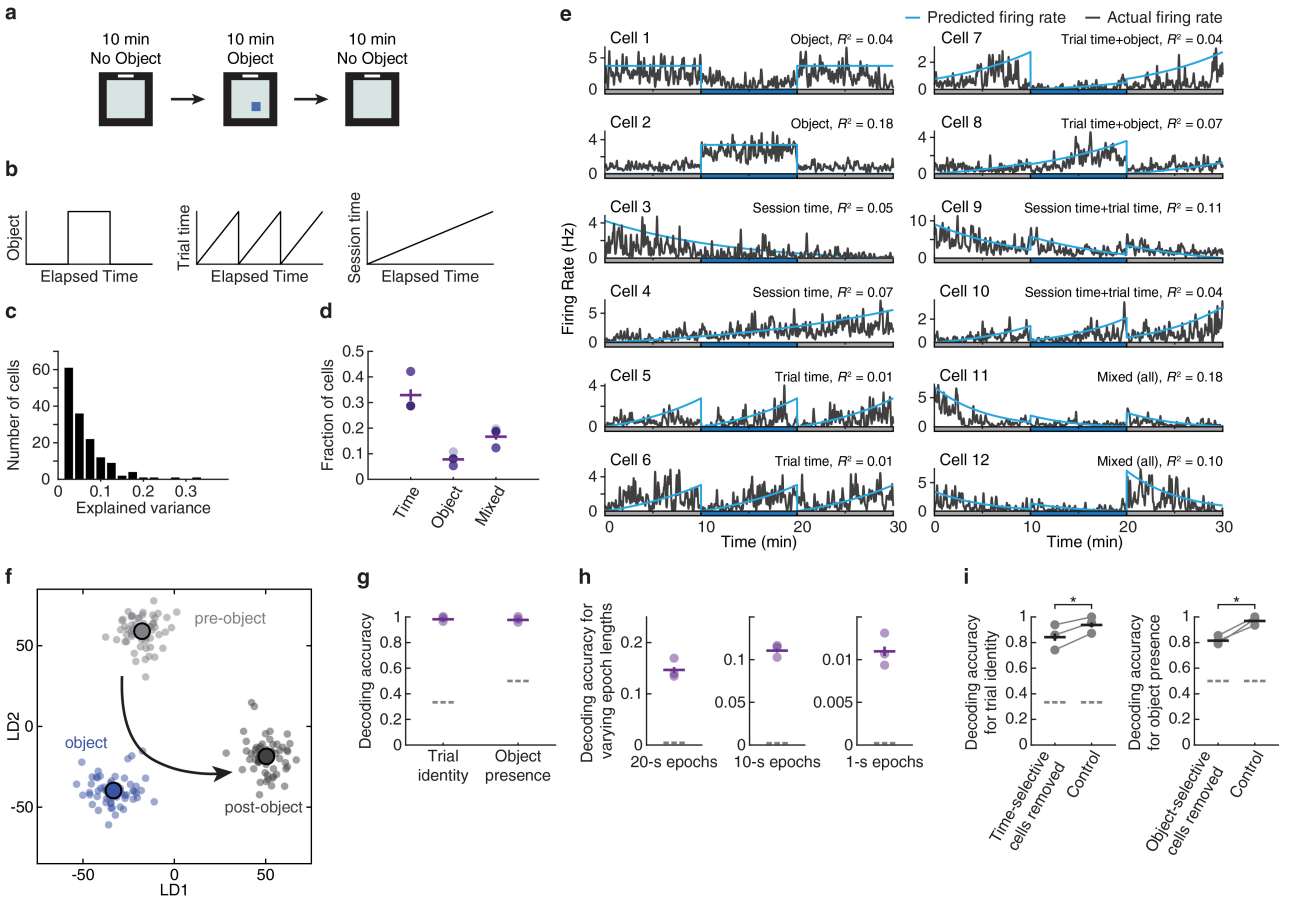
Extended Data Fig. 4 | Decoding of temporal epochs. a, 2D projections of neural population responses during the BW12 experiment for trial periods only. Axes correspond to the first two linear discriminants (LD1 and LD2; arbitrary units). Left column shows LEC population responses, middle column shows CA3 population responses, right column shows MEC population responses, each from an example animal. The wall colour of each trial is indicated by a shade of green (black walls) or purple (white walls), with progression of shade from dark to light indicating the progression of trials. **b**, Fraction of variance explained by the first 20 principal components for each area. Principal components were computed using PCA on raw data. Lines indicate variance explained for individual animals ($n = 3$, 3 and 2 animals for LEC, CA3 and MEC, respectively). **c**, Regressing the first two principal components from PCA results for individual animals against time leads to significant fits for all areas, but substantially higher explained variance for LEC. $P < 0.001$ (LEC versus CA3, $t(4) = 9.79$), $P < 0.01$ (LEC versus MEC, $t(3) = 6.13$), unpaired t-test. Top row, example fits for individual animals are shown with black lines indicating time, coloured lines indicating regression fit and R^2 values indicated for the example fit. Bottom two rows, first two principal components for example fits. **d**, Illustration of cross-validation procedure: fivefold cross-validation is shown for data containing four temporal epochs, with five time bins in each epoch. A different subset of time bins is used as test data for each iteration of the cross-validation procedure. Actual data consisted of 24 epochs (trial and intertrial periods) with 25 or 14 time bins in each epoch, and tenfold cross-validation was used. **e**, Z-scored decoding accuracy using cells recorded in a single day (left, $P = 0.68$, one-sided binomial test, $n = 46$ days), pairs of consecutive days (middle, $P = 0.29$, one-sided binomial test, $n = 72$ pairs), pairs of days separated by half the total number of recording days (right, $P = 0.37$, one-sided binomial test, $n = 44$ pairs). **f**, Decoding accuracy for temporal epoch using behaviour tracking data in place of neural activity ($n = 3$ animals). 'All' tracking data consisted of the animal's position, velocity, acceleration and head direction. **g**, Left, decoding accuracy for temporal epoch using BW4 data, compared to decoding accuracy using BW12 data. $P = 0.32$ ($t(5) = 1.09$), unpaired t-test; matched population size = 47 cells, $n = 4$ and 3 animals for BW4 and BW12 respectively. Right, decoding accuracy for wall colour from the BW4 experiment, compared to decoding accuracy using matched data from BW12 experiment. $P = 0.35$ ($t(5) = 1.03$), unpaired t-test; matched population size = 47 cells, $n = 4$ and 3 animals for BW4 and BW12 respectively. **h**, Decoding accuracy for temporal epoch using BW4 data from the LEC, MEC, CA3, CA2 and CA1. One-way ANOVA, $F(4) = 20.78$, $P < 1 \times 10^{-5}$, post hoc Bonferroni multiple comparisons test, $P < 0.005$ (for each comparison against LEC); matched population size = 24 cells, $n = 7$, 2, 7, 3 and 3 animals for LEC, MEC, CA3, CA2 and CA1 respectively. **i**, Relation between population size and decoding accuracy for the LEC, CA3 and MEC. Left, decoding accuracy for varying population sizes; each line indicates the curve fit to data (shown as points) from one animal, with colours indicating recording area ($n = 3$, 3 and 2 animals for LEC, CA3 and MEC, respectively). Right, relation between population size and decoding accuracy for the LEC and CA3, pooled across animals from BW12 and BW4 experiments (data pooled from $n = 7$ animals for both LEC and CA3). **j**, Decoding accuracy for wall colour from trial period activity alone for the LEC, CA3 and MEC. $P = 0.47$ (LEC versus CA3, $t(4) = 0.81$) unpaired t-test; matched population size = 28 cells, $n = 3$, 3 and 2 animals for LEC, CA3 and MEC, respectively. **k**, Decoding accuracy for wall colour using data that was shuffled in time. $P = 0.74$ ($t(2) = 0.38$), two-tailed paired t-test; $n = 3$ animals. **l**, Decoding accuracy for wall colour using a subpopulation with all wall-colour-selective cells removed, compared to size-matched populations that were randomly drawn from full population. $P = 0.12$ ($t(2) = 2.66$), paired t-test; $n = 3$ animals with 77, 126 and 195 cells not selective for wall colour. For **f**, **g**, **j**–**l**, circles indicate individual animals, solid lines indicate mean decoding accuracy ± s.e.m., dashed lines indicate chance levels.



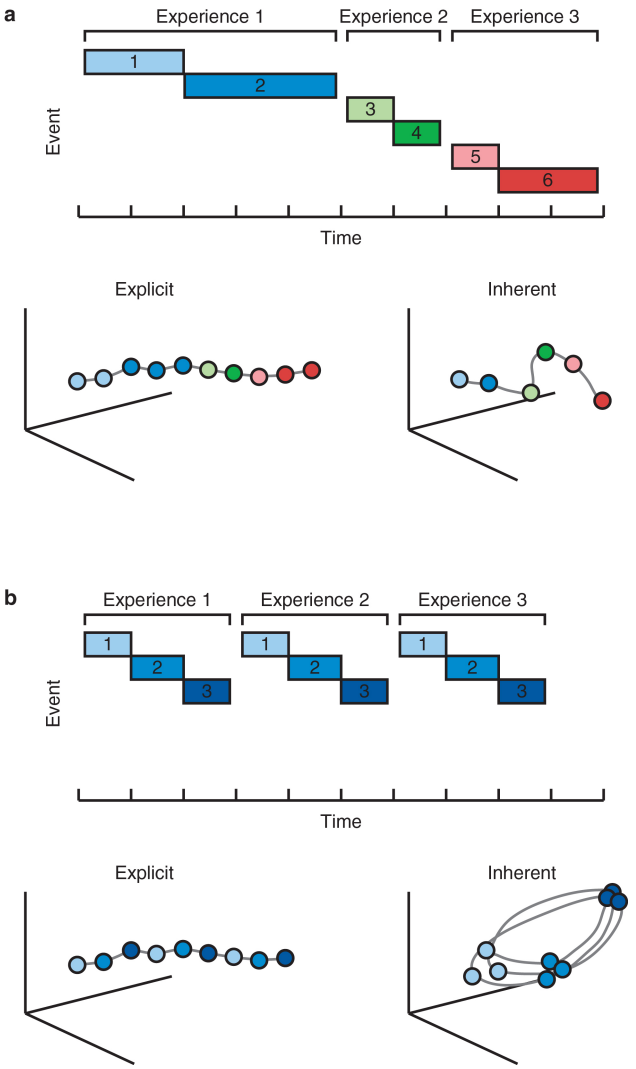
Extended Data Fig. 5 | Evolution of LEC population dynamics. Structure and evolution of LEC (purple, $n = 3$ animals), CA3 (blue, $n = 3$ animals) or MEC (gold, $n = 2$ animals) activity, quantified by applying several distance measurements to population activity states (see Supplementary Information). The distance being measured is illustrated in cartoon form above each plot (d_1: distance between time bin 1 and time bin 2, d_2: distance between time bin 2 and time bin 3, and so on). **a**, Manhattan distance between the population state for the middle time bin of a trial or intertrial period and all other time bins within that period. **b**, Manhattan distance between the population state for the last time bin of the preceding trial/intertrial period and all time bins within the current intertrial/trial period. **c**, Left, Manhattan distance between the population state for the first time bin of a trial or intertrial period and all subsequent time bins within that period. Right, fitting distance using an autoregressive (AR) model. Red line indicates model fit. **d**, Manhattan distance between consecutive pairs of population states across trial or intertrial periods. **e**, Pairwise angles, measured across consecutive points in time along neural trajectories during trial or intertrial periods as the angle between two vectors, each defined as the difference of consecutive population states. **f**, Manhattan distance between the overall mean population state for the first trial or intertrial period and the overall mean population states of all subsequent trial or intertrial periods. Solid lines indicate mean distance across all animals, shaded area indicates s.e.m., distance measures were all z-scored.



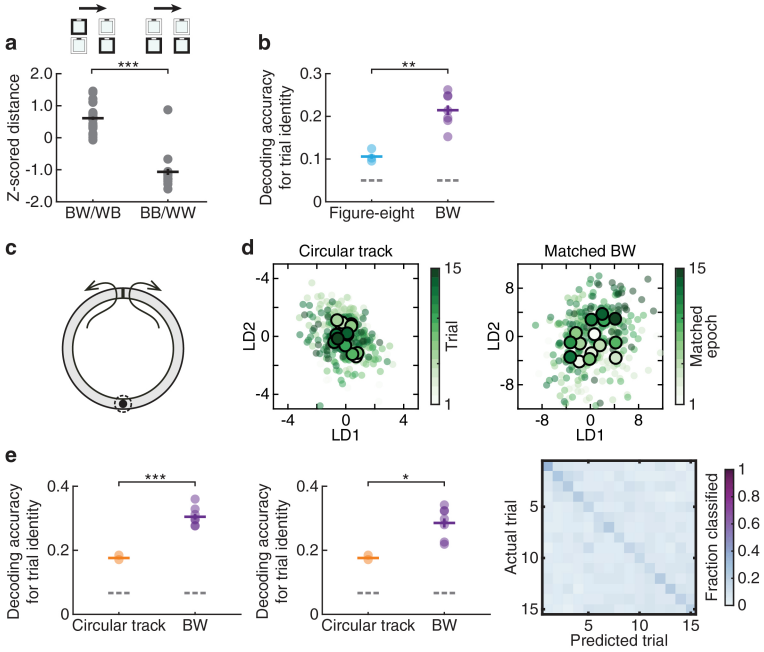
Extended Data Fig. 6 | Decoding shortened temporal epochs. **a**, Decoding accuracies for different temporal epoch lengths across the LEC, MEC, CA3, CA2 and CA1 using BW4 data ($n = 7, 2, 7, 3$ and 3 animals for LEC, MEC, CA3, CA2 and CA1, respectively). Decoding accuracy for 20-s epochs using LEC data was significantly better than all hippocampal areas. $P < 1 \times 10^{-5}$ ($F(17) = 20.69$) one-way ANOVA, post hoc Bonferroni multiple comparisons test; $P < 0.05$ (LEC versus MEC), $P < 0.001$ (LEC versus CA3, CA2 or CA1); matched population size = 24 cells. Decoding accuracy was higher for the LEC and MEC than for the CA3 for 10-s epochs. $P < 0.001$ (LEC versus CA3, all other comparisons were not significant, $F(17) = 8.00$), one-way ANOVA, post hoc Bonferroni multiple comparisons test; matched population size = 24 cells. Decoding accuracy was similar across all areas for 1-s epochs. $P = 0.05$ ($F(17) = 2.95$), one-way ANOVA. **b**, Confusion matrix from an example LEC animal for 10-s epochs. The matrix contains 468 epochs (each trial period of 250 s divided into 25 10-s epochs, each intertrial period of 140 s divided into 14 10-s epochs, 39 epochs per trial/intertrial pair, 12 total pairs across the session, giving 468 epochs). Left, confusion matrix for the entire session. Right, sum across trial/intertrial sections of the whole session matrix (outlined by dashed line in whole session matrix). **c**, Decoding accuracy for re-binned confusion matrices, see Supplementary Information (circles and solid lines with black indicates mean, shade indicates the same individual; $P < 1 \times 10^{-4}$ (comparing across epoch lengths, $F(3) = 32.28$) one-way ANOVA; $P = 0.12$ (trial versus 20 s), $P < 0.001$ (trial versus 10 s, and trial versus 1 s)) compared to decoding accuracy for re-binned confusion matrices following shuffling along columns, with diagonals preserved (triangles and dashed lines, with black indicating mean and shade indicating individual animals; comparing decoding accuracies from shuffled and unshuffled confusion matrices: two sided paired t-tests, for 20 s: $t(2) = 7.42$, $P < 0.05$; for 10 s: $t(2) = 14.31$, $P < 0.01$; for 1 s: $t(2) = 4.46$, $P < 0.05$). Grey dashed line indicates chance at 4.2%, shade indicates individual animals. **d**, Decoding accuracy for different temporal epoch lengths for the LEC (matched population size = 90 cells, $n = 3$ animals), based only on activity within individual trial or intertrial periods (top or bottom, respectively). Decoding accuracy was measured for each trial or intertrial period individually, and then the average was taken across all trials or intertrial periods, respectively, for each animal. Temporal bin size was: 1 s for 20-s-long epochs, 500 ms for 10-s-long epochs, and 50 ms for 1-s-long epochs. Chance levels for trial periods were 8.3%, 4.0% and 0.4%, and 14.3%, 7.1% and 0.7% for intertrial periods. **e**, Confusion matrices from an example LEC animal for single-trial or intertrial periods as in **d**, using 20-s (top) and 10-s (bottom) epochs, with intertrial results on the left and trial results on the right. For **a** and **d**, circles indicate individual animals, solid lines indicate mean decoding accuracies ± s.e.m., dashed lines indicate chance levels.



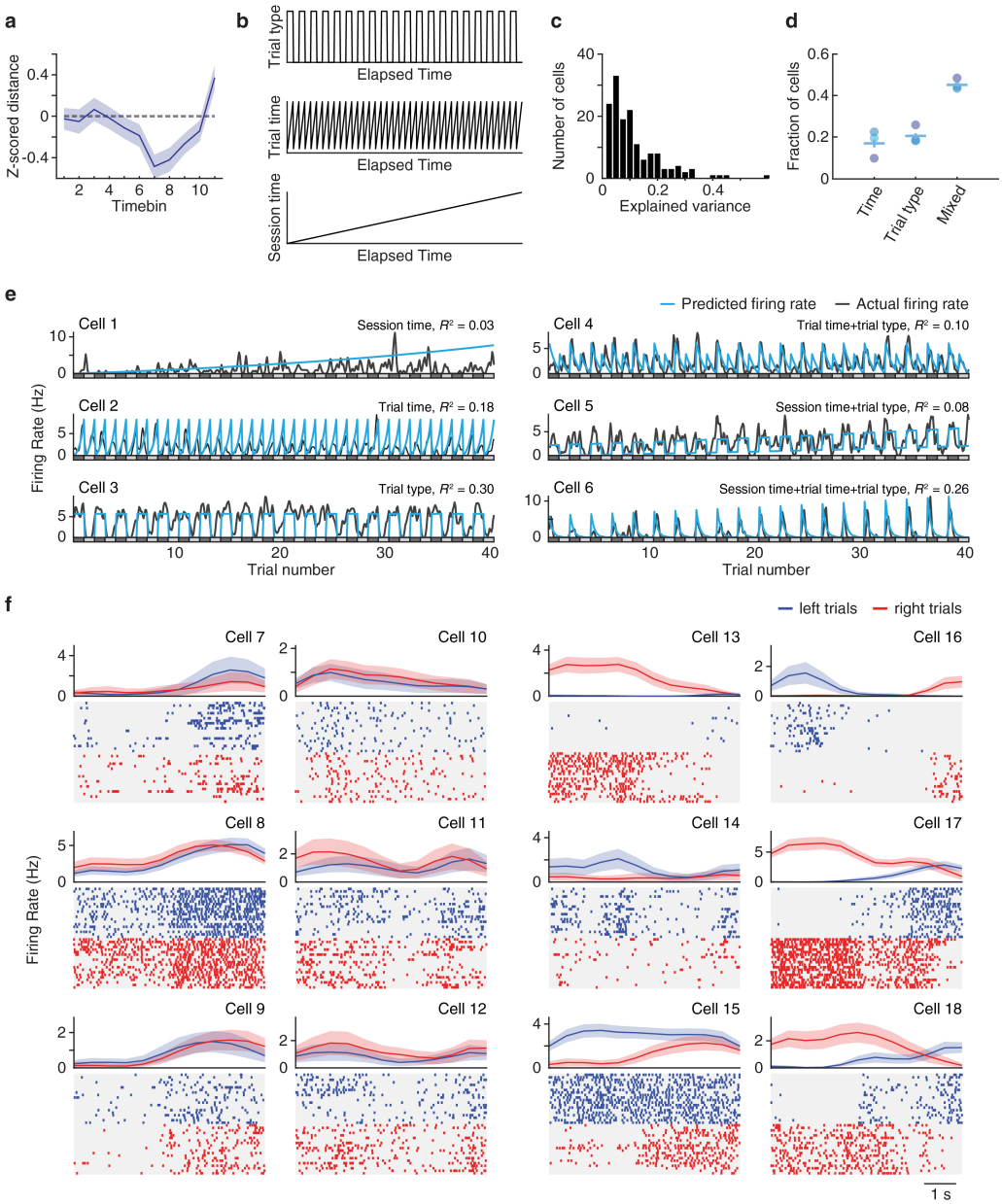
Extended Data Fig. 7 | Temporal coding within a fixed environmental context. **a**, Experimental design: 10 min without object, 10 min with object, and 10 min without object. **b**, Predictors used for object GLM: object presence, trial time and session time (24.0% selective for session time, 4.8% selective for trial time, 4.3% selective for a mixture of trial time and session time, percentages averaged across all animals). **c**, Explained variance for all LEC cells fit by GLM for object experiments ($n = 150$ cells). Average explained variance was 0.05. **d**, Distribution of selectivity for time, object and mixtures of time and object for single LEC cells ($n = 3$ animals with 56, 57 and 150 cells). Shade indicates the same individual across the different variables. **e**, Examples of GLM fit results for 12 cells from the object experiment with selectivity for different features. The firing rate of each cell is shown in grey, with the model-predicted firing rate in blue. The R^2 value is shown for each cell. **f**, 2D projection of LEC neural population responses during object experiment from example animal. **g**, Decoding accuracy for trial identity or object presence ($n = 3$ animals). **h**, Decoding accuracies for temporal epochs of shortened length ($n = 3$ animals, matched population size = 56 cells). **i**, Decoding accuracy for trial identity (left) or object presence (right), with cells selective for decoded variable removed, compared to size-matched populations randomly drawn from full population. Trial identity: $P < 0.05$ ($t(2) = 4.95$), paired t-test; $n = 29$, 26 and 79 cells not selective for time; object presence: $P < 0.05$ ($t(2) = 7.17$), paired t-test; $n = 39$, 47, and 110 cells not selective for object presence. Circles indicate individual animals, solid lines indicate mean ± s.e.m. of described measurement, dashed lines indicate chance levels.



Extended Data Fig. 8 | Explicit versus inherent mechanisms for temporal coding. a, Top, a series of experiences occurs, each containing different event content and spanning different amounts of time. Bottom, two different ways in which temporal information within this series of experiences may be encoded. For both cases, a population code is used, but this may just as easily be replaced by a rate code within single cells. An explicit mechanism (left) purposely represents the passage of time, such that each chunk of time is represented equally. Thus, two experiences with the same temporal length but differing numbers of events would correspond to the same change in activity. An inherent mechanism (right) encodes temporal information entirely by representing the events within each experience. Thus, two experiences with the same temporal length but differing numbers of events would correspond to differing changes in activity. In either case, the high dimensionality of the representations would allow temporal information to be read out easily by downstream readout neurons, for example, cells in the hippocampus. **b**, As in **a**, but in this example, instead of a series of different experiences, the same experience is repeated three times (analogous to performing a learned task three times). Here, an explicit mechanism for temporal coding would exhibit the same amount of change in activity as in **a**, whereas an inherent mechanism would exhibit considerably reduced differences in activity across the experiences.



Extended Data Fig. 9 | Decoding trial identity with additional matched data. a, Euclidean distance between mean LEC population states for pairs of adjacent trials with different or same wall colour. $P < 10^{-9}$ ($t(31) = 8.81$), unpaired t-test; $n = 12$ and 21 for same colour (BB/WW) and different colour (BW/WB) transitions, respectively, pooled from three animals. **b**, Decoding accuracy for trial identity during the figure-eight experiment, compared to decoding accuracy for temporal epoch using matched data from BW experiments that did not account for intertrial intervals or trial type. $P < 0.01$ ($t(8) = 4.91$), unpaired t-test; matched population size = 31 cells, $n = 3$ and 7 animals for figure-eight and BW, respectively. **c**, Circular-track task, in which animals alternated between clockwise and anticlockwise runs for 15 consecutive back and forth laps. Black circle indicates midpoint of the track. **d**, Left, 2D projection of the LEC neural population response during circular track experiment from example animal. Right, 2D projection of the LEC neural population response during matched periods from BW experiments. **e**, Left, decoding accuracy for trial identity during circular track experiment compared to decoding accuracy for temporal epoch using matched data from BW experiments. $P < 0.0001$ ($t(7) = 6.21$), unpaired t-test; matched population size = 47 cells, $n = 2$ and 7 animals for circular track and BW, respectively. Right, as in the left panel, but using temporally consecutive data that did not account for intertrial intervals or running direction. $P < 0.05$ ($t(7) = 3.17$), unpaired t-test; matched population size = 47 cells, $n = 2$ and 7 animals for circular track and BW, respectively. **f**, Confusion matrix for decoding trial identity in the circular track experiment from example animal. Circles indicate individual animals, solid lines indicate mean ± s.e.m. of described measurement, dashed lines indicate chance levels.



Extended Data Fig. 10 | Additional characterization of LEC activity during figure-eight task. a, Manhattan distance between consecutive pairs of population states (binned in 500-ms time bins) across single trials, averaged across all animals. **b**, Predictors used for the figure-eight GLM: trial type, trial time and session time (1.3% selective for session time, 9.3% selective for trial time, 6.5% selective for a mixture of trial time and session time, percentages averaged across all animals). **c**, Explained variance for all LEC cells fit by GLM for the figure-eight experiment ($n = 149$ cells). Average explained variance was 0.10. **d**, Distribution of selectivity for time, trial type and a mixture of time and trial type for single LEC cells, determined using a GLM ($n = 3$ animals with 72, 76 and 31 cells). Circles indicate individual animals, solid lines indicate mean fraction of cells ± s.e.m., shade indicates the same individual across the different variables. **e**, Examples of GLM fit results for six cells from the figure-eight experiment with selectivity for different features. The firing rate of each cell is shown in grey, with the model-predicted firing rate in blue. R^2 value is shown for each cell. **f**, LEC activity for 12 example cells during the figure-eight task. Each plot shows the mean firing rate (top, 95% percentile confidence interval shaded), and peristimulus time histograms for left-turn (middle) and right-turn trials (bottom), with time centred on the point at which the animal reaches the base of the central stem on each trial. Cells 7–12 exhibited similar firing patterns for both left- and right-turn trials, including during the first 3 s of the trial, in which the animal occupied a different spatial location for left- versus right-turn trials. Such activity may be used for temporal information. Cells 13–18 exhibited highly divergent firing patterns for left- and right-turn trials, which may reflect the animal's spatial location, the behavioural context of the trial, or a combination of the two variables. All example cells exhibited relatively stable firing across trials, a common feature observed during the figure-eight task.

ARTICLE

https://doi.org/10.1038/s41586-018-0466-7

Role of glutamine synthetase in angiogenesis beyond glutamine synthesis

Guy Eelen[1,2,3,21]*, Charlotte Dubois[1,3,21], Anna Rita Cantelmo[1,3,15], Jermaine Goveia[1,3], Ulrike Brüning[1,3,16], Michael DeRan[4], Gopala Jarugumilli[4], Jos van Rijssel[5], Giorgio Saladino[6], Federico Comitani[6], Annalisa Zecchin[1,3], Susana Rocha[7], Rongyuan Chen[2], Hongling Huang[1,3,17], Saar Vandekeere[1,3], Joanna Kalucka[1,3], Christian Lange[1,3,18], Francisco Morales-Rodriguez[1,3], Bert Cruys[1,3], Lucas Treps[1,3], Leanne Ramer[1,3,19], Stefan Vinckier[1,3], Katleen Brepoels[1,3], Sabine Wyns[1,3], Joris Souffreau[1,3], Luc Schoonjans[1,3], Wouter H. Lamers[8], Yi Wu[9], Jurgen Haustraete[10,11], Johan Hofkens[7], Sandra Liekens[12], Richard Cubbon[1,3,20], Bart Ghesquière[13], Mieke Dewerchin[1,3], Francesco L. Gervasio[6,14], Xuri Li[2]*, Jaap D. van Buul[5], Xu Wu[4] & Peter Carmeliet[1,2,3]*

Glutamine synthetase, encoded by the gene *GLUL*, is an enzyme that converts glutamate and ammonia to glutamine. It is expressed by endothelial cells, but surprisingly shows negligible glutamine-synthesizing activity in these cells at physiological glutamine levels. Here we show in mice that genetic deletion of *Glul* in endothelial cells impairs vessel sprouting during vascular development, whereas pharmacological blockade of glutamine synthetase suppresses angiogenesis in ocular and inflammatory skin disease while only minimally affecting healthy adult quiescent endothelial cells. This relies on the inhibition of endothelial cell migration but not proliferation. Mechanistically we show that in human umbilical vein endothelial cells *GLUL* knockdown reduces membrane localization and activation of the GTPase RHOJ while activating other Rho GTPases and Rho kinase, thereby inducing actin stress fibres and impeding endothelial cell motility. Inhibition of Rho kinase rescues the defect in endothelial cell migration that is induced by *GLUL* knockdown. Notably, glutamine synthetase palmitoylates itself and interacts with RHOJ to sustain RHOJ palmitoylation, membrane localization and activation. These findings reveal that, in addition to the known formation of glutamine, the enzyme glutamine synthetase shows unknown activity in endothelial cell migration during pathological angiogenesis through RHOJ palmitoylation.

Endothelial cells (ECs) line the lumen of blood vessels. Emerging evidence reveals that EC metabolism controls vessel sprouting (angiogenesis)[1–3]. Although glutamine catabolism in ECs has been recently characterized[4], it is not known whether glutamine anabolism controls angiogenesis in vivo. Glutamine is a carbon and nitrogen donor for the production of biomolecules and is involved in redox homeostasis. Most cells take up glutamine, and therefore do not need to synthesize it. However, certain cell types express *GLUL* (glutamate-ammonia ligase), the gene that encodes the enzyme glutamine synthetase (GS), which is capable of de novo glutamine production from glutamate and ammonia in a reaction that requires ATP and Mg^{2+} or Mn^{2+}. GS also serves another biochemical function—the clearance of ammonia—but this is best described for hepatocytes, astrocytes and muscle. ECs also express GS[5], although its role and importance in angiogenesis is unclear given that ECs are exposed to high plasma glutamine levels. Global deficiency of GS causes embryonic lethality, presumably owing to the inability to detoxify ammonia[6]. GS deficiency in humans is extremely rare and leads to multi-organ failure and infant death[7]. Whether and how GS affects angiogenesis has, to our knowledge, not yet been analysed. Here we characterized the role and importance of GS in vessel sprouting.

Vessel sprouting requires endothelial GS

We checked the expression of GS in endothelial cells of the retinal microvasculature with a genetic *Glul* reporter mouse (*Glul*$^{+/GFP}$ mice with a nucleus-targeted GFP–lamin A fusion reporter transgene in the *Glul* open reading frame of one allele[6]). GFP tracing in the post-natal day (P)5 retinal plexus, co-stained with the endothelial cell marker isolectin B4 (IB4; red), revealed endothelial expression of GFP (and therefore expression of GS) in the microvasculature (Fig. 1a).

Human umbilical vein ECs (HUVECs) expressed *GLUL* at similar levels to human colon ECs, liver ECs, human umbilical artery ECs and blood outgrowth ECs, but at a lower level than lung ECs (Extended Data Fig. 1a). However, the expression of GS in ECs or isolated mouse liver ECs was lower than in HEPG2 hepatocellular carcinoma cells or astrocytes (Extended Data Fig. 1a–c), which are known to highly express GS. Glutamine withdrawal (below the physiological concentration of 0.6 mM) increased GS protein levels in ECs

[1]Center for Cancer Biology, University of Leuven, Leuven, Belgium. [2]State Key Laboratory of Ophthalmology, Zhongshan Ophthalmic Center, Sun Yat-Sen University, Guangzhou, China. [3]Center for Cancer Biology, VIB, Leuven, Belgium. [4]Cutaneous Biology Research Center, Massachusetts General Hospital, Harvard Medical School, Charlestown, MA, USA. [5]Department of Molecular Cell Biology, Sanquin Research and Landsteiner Laboratory, Academic Medical Centre, University of Amsterdam, Amsterdam, The Netherlands. [6]Department of Chemistry, University College London, London, UK. [7]Molecular Imaging and Photonics, University of Leuven, Leuven, Belgium. [8]Tytgat Institute for Liver and Gastrointestinal Research, Academic Medical Center, University of Amsterdam, Amsterdam, The Netherlands. [9]Center for Cell Analyses and Modelling, University of Connecticut Health Centre, Farmington, CT, USA. [10]Inflammation Research Center, VIB, Ghent, Belgium. [11]Department of Biomedical Molecular Biology, Ghent University, Ghent, Belgium. [12]Department of Microbiology and Immunology, University of Leuven, Leuven, Belgium. [13]Metabolomics Core Facility, Center for Cancer Biology, VIB, Leuven, Belgium. [14]Institute of Structural Molecular Biology, University College London, London, UK. [15]Present address: Université de Lille, INSERM U1003, Physiologie Cellulaire, Lille, France. [16]Present address: Max-Delbrück-Center for Molecular Medicine, Berlin, Germany. [17]Present address: Immunology Department, St. Jude Children's Research Hospital, Memphis, TN, USA. [18]Present address: DFG-Research Center for Regenerative Therapies, Technical University Dresden, Dresden, Germany. [19]Present address: Department of Biomedical Physiology and Kinesiology, Simon Fraser University, Burnaby, British Columbia, Canada. [20]Present address: Division of Cardiovascular and Diabetes Research, Multidisciplinary Cardiovascular Research Centre, University of Leeds, Leeds, UK. [21]These authors contributed equally: Guy Eelen, Charlotte Dubois. *e-mail: guy.eelen@kuleuven.vib.be; lixr6@mail.sysu.edu.cn; peter.carmeliet@kuleuven.vib.be

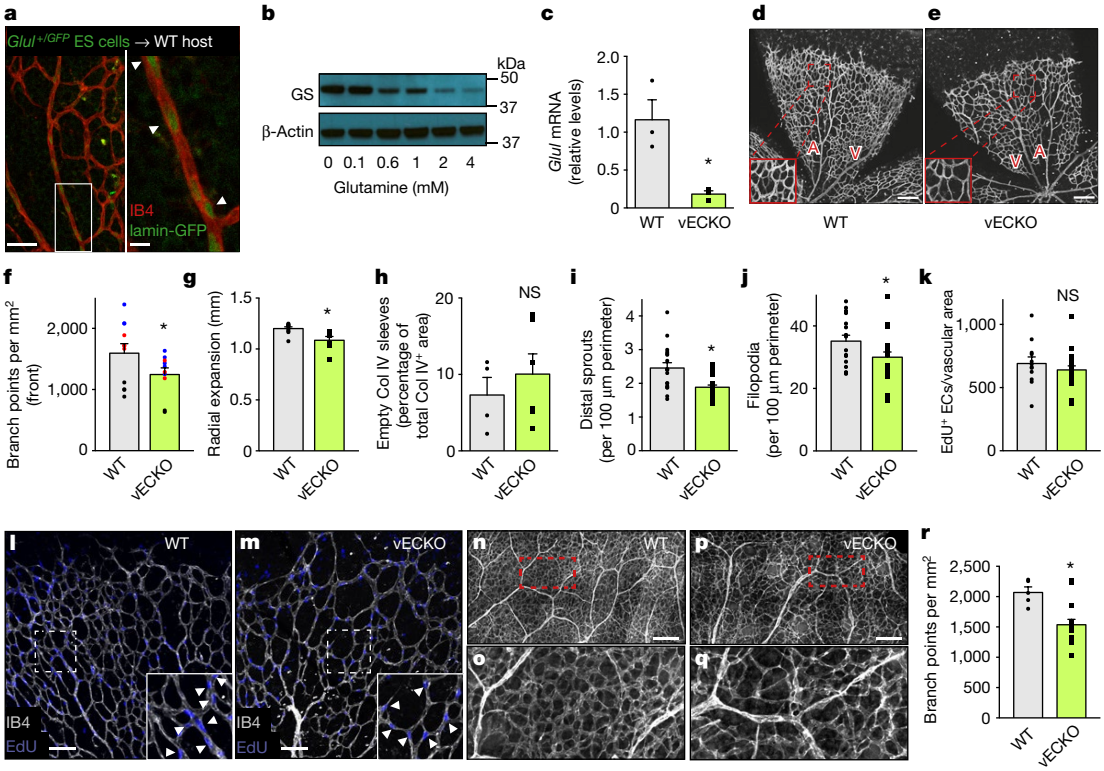


Fig. 1 | EC-specific deletion of GS causes vascular defects in vivo. **a**, GS expression (arrowheads) in the retinal microvasculature (co-stained with IB4) of P5 chimaeric *Glul*$^{+/GFP}$ pups. The area indicated by the white box is shown magnified on the right. ES cells, embryonic stem cells. **b**, GS protein levels in HUVECs under different concentrations of extracellular glutamine. **c**, *Glul* mRNA levels upon activation of *VE-cadherin-cre*^{ERT2}. **d**–**g**, IB4 staining of P5 retinal vascular plexuses from wild-type (WT) (**d**) and *Glul*^{vECKO} (**e**) mice (with magnifications shown in the insets; A, artery; V, vein) and quantification of branch points at the front of the plexus (**f**) and radial expansion of the plexus (**g**). **h**, Vessel regression (area of collagen IV (Col IV)$^{+}$IB4^{-} vessel sleeves as a percentage of total Col IV^{+} area) in retinas from P5 wild-type and *Glul*^{vECKO} pups. **i**, **j**, Quantification of distal sprouts (**i**) and filopodia (**j**) at the retinal vascular front. **k**–**m**, IB4 (grey)/EdU (cyan) double staining of P5 wild-type (**l**) and *Glul*^{vECKO} (**m**) retinas (magnifications are shown in the insets, with arrowheads denoting EdU^{+} ECs) and quantification (**k**) of EdU^{+} ECs at the front of the plexus. **n**–**r**, CD31-stained dermal dorsal blood vasculature in E16.5 wild-type (**n**, **o**) and *Glul*^{vECKO} (**p**, **q**) mice with boxed regions magnified in **o** and **q**, and quantification of the number of branch points (**r**). All data are mean ± s.e.m., with individual data points shown in **c**, **f**–**k** and **r**; $n = 2$ individual experiments (**a**, **b**); values of n (individual mice) for wild-type and *Glul*^{vECKO}, respectively, are: 3 and 3 (**c**); 11 and 10 (**f**); 10 and 7 (**g**); 4 and 6 (**h**); 18 and 22 (**i**); 17 and 21 (**j**); 12 and 22 (**k**); 5 and 15 (**r**), from 2 (**g**, **h**, **r**), 3 (**f**), 4 (**k**) or 5 (**i**, **j**) litters. NS, $P > 0.05$, $^{*}P < 0.05$ according to Student's t-test (**c**, **g**, **h**, **i**, **j**, **k**, **r**) or mixed-models R statistics (**f**). Exact P values are as follows: 0.0215 (**c**); 0.0141 (**f**); 0.0063 (**g**); 0.4902 (**h**); 0.0009 (**i**); 0.0484 (**j**); 0.3837 (**k**); 0.0046 (**r**). Scale bars: 10 μm (**a** right), 50 μm (**a** left), 100 μm (**l**, **m**), 200 μm (**d**, **e**, **n**, **p**). For gel source images, see Supplementary Fig. 1.

(Fig. 1b, Extended Data Fig. 1b), as has been previously documented for other cell types[8].

We intercrossed *Glul*$^{lox/lox}$ mice with two different EC-specific tamoxifen-inducible Cre driver lines—*VE-cadherin(PAC)-cre*^{ERT2} and *Pdgfb-cre*^{ERT2} mice—to obtain *Glul*^{vECKO} and *Glul*^{pECKO} mice, respectively (in which ECKO indicates endothelial cell knockout; VE-cadherin is also known as *Cdh5*). Correct recombination of the excised *Glul* allele was confirmed (Extended Data Fig. 1d, e) and caused an average reduction of 84% in *Glul* mRNA levels in mouse liver ECs isolated from *Glul*^{vECKO} mice (Fig. 1c). In the neonatal retina, vascular plexuses in P5 *Glul*^{vECKO} mice showed hypobranching and reduced radial expansion, whereas vessel coverage by NG2^{+} pericytes (NG2, chondroitin sulfate proteoglycan 4) and vessel regression (number of empty collagen IV^{+} sleeves) were unaffected (Fig. 1d–h, Extended Data Fig. 1f, g). However, the numbers of filopodia at the vascular front and of distal sprouts with filopodia, both parameters of EC migration, were lower in *Glul*^{vECKO} mice (Fig. 1i, j). Furthermore, the complexity of the vasculature at the utmost leading front of the plexus was decreased, as determined by counting the number of branches in distal sprouts (Extended Data Fig. 1h). By contrast, quantification of IB4^{+}EdU^{+} cells (EdU, 5-ethynyl-2′-deoxyuridine) revealed no difference in the number of proliferating ECs (Fig. 1k–m, Extended Data Fig. 1i). Hypobranching was also observed in the dorsal dermal blood vasculature in embryonic day (E)16.5 *Glul*^{vECKO} embryos (Fig. 1n–r). A similar retinal phenotype was observed in *Glul*^{pECKO} mice (Extended Data Fig. 1j–m). Therefore, loss of endothelial GS causes vascular defects by impairing EC migration but not proliferation.

The retinal vascular defect was restored over time (Extended Data Fig. 1n–u), and at six weeks, *Glul*^{vECKO} mice (with *Glul* deleted in ECs at P1–P3) did not show overt vascular defects (Extended Data Fig. 1v–ag). *Glul*^{vECKO} mice gained normal body weight, and blood biochemistry and haematological profiles were normal at six weeks (Extended Data Table 1). Vascular restoration may relate to the possibility that homozygous mutant ECs were outcompeted over time by residual wild-type ECs, in which recombination did not occur (as documented in mice with endothelial knockout of other key metabolic genes[9]) or because of other compensatory adaptations. Alternatively, the results raise the question of whether the effect of endothelial GS loss may be larger in growing (motile) ECs during vascular development than in quiescent (non-motile) ECs during adulthood in healthy conditions.

We then explored whether pharmacological blockade of GS with methionine sulfoximine (MSO), which irreversibly blocks the catalytic activity of GS, reduced pathological angiogenesis. First, in the oxygen-induced model of retinopathy of prematurity (ROP)[2,3], the treatment of pups with MSO reduced the formation of pathological vascular tufts (Fig. 2a–c), while modestly increasing the vaso-obliterated area (Fig. 2d, Extended Data Fig. 1ah, ai). Second, we used the corneal micro-pocket assay, with slow-release pellets containing basic

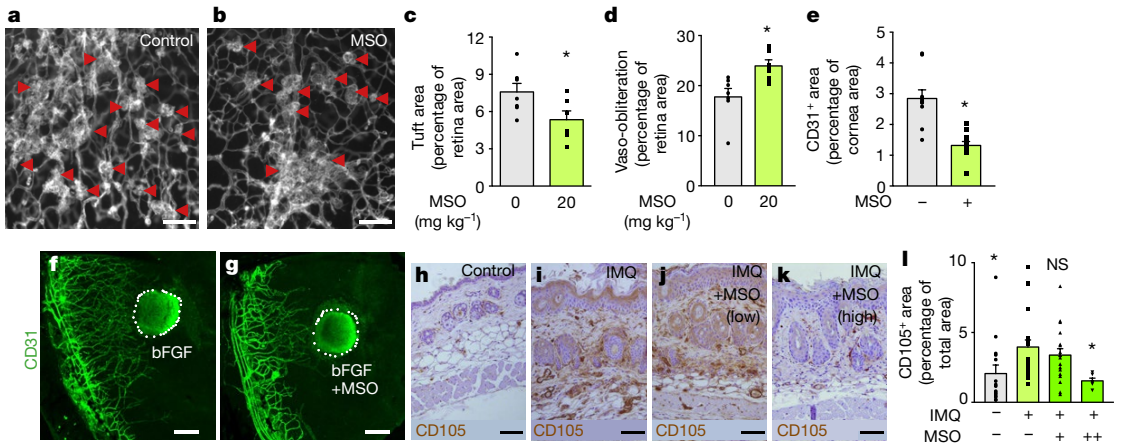


Fig. 2 | GS inhibition mitigates pathological angiogenesis. a–d, Retinal flat-mounts of ROP mice treated with vehicle (**a**) or 20 mg kg^{-1} day^{-1} MSO (**b**). Quantification of vascular tuft (**c**) and vaso-obliterated area (**d**) in control and MSO-treated ROP mice. **e–g**, Quantification (**e**) of $CD31^+$ (green) neo-vessels in corneal flat-mounts from mice in corneal pocket assays with bFGF pellets (demarcated by a dotted white line) with vehicle (**f**) or MSO (**g**). **h–l**, CD105 staining of untreated skin (**h**), IMQ-treated skin (**i**), IMQ + low-dose MSO-treated skin (**j**), IMQ + high-dose MSO-treated skin (**k**) and quantification of $CD105^+$ area (**l**). All data are mean ± s.e.m., with individual data points shown for **c**, **d**, **e** and **l**; values of *n* (individual mice) for control and MSO-treated are: 7 and 6 (**c**, **d**), 10 and 11 (**e**) from 3 litters (**c**, **d**) and 2 experiments (**e**). In **l**, $n = 15$ mice for control, $n = 22$ for IMQ, $n = 18$ for IMQ + MSO low-dose (indicated by +), and $n = 6$ for IMQ + MSO high-dose (indicated by ++) from 3 experiments. NS, $P > 0.05$, $^*P < 0.05$ according to Student's *t*-test (**c**, **d**, **e**) or one-way analysis of variance (ANOVA) with Dunnett's multiple comparisons versus IMQ (**l**). Exact *P* values are as follows: 0.0459 (**c**); 0.0145 (**d**); <0.0001 (**e**); control versus IMQ: 0.0278; MSO low versus IMQ: 0.7283; MSO high versus IMQ: 0.0451 (**l**). Scale bars: 100 μm (**a**, **b**), 200 μm (**f**, **g**), 75 μm (**h–k**).

fibroblast growth factor (bFGF) as a mouse model of corneal neovascularization. Inclusion of MSO in the pellet reduced the formation of new $CD31^+$ blood vessels in the otherwise avascular cornea (Fig. 2e–g). Finally, we used an imiquimod (IMQ)-based mouse model of inflammation-driven skin psoriasis, and found a marked dose-dependent reduction of the $CD105^+$ EC area upon topical treatment of the affected skin with MSO (Fig. 2h–l). Therefore, pharmacological GS blockade inhibits pathological angiogenesis in the inflamed skin and in several eye disorders.

Silencing *GLUL* reduces EC migration

We then used *GLUL* knockdown ($GLUL^{KD}$) HUVECs (mediated by short hairpin RNA (shRNA); >80% silencing; Extended Data Fig. 2a) in in vitro spheroid-sprouting assays to assess vessel sprouting. $GLUL^{KD}$ ECs showed a reduced number of sprouts per spheroid and a reduction in the total sprout length (Fig. 3a, b, e, f). Re-introduction of a shRNA-resistant *GLUL* ($rGLUL^{OE}$; in which r indicates shRNA-resistance and OE indicates overexpression) rescued the sprouting defect (Extended Data Fig. 2b, c). The sprouting defect in $GLUL^{KD}$ spheroids was maintained upon mitotic inactivation of ECs with mitomycin C (Fig. 3c–f), which further suggests a defect in EC motility. In agreement with this, at physiological glutamine levels *GLUL* knockdown did not affect EC proliferation (Fig. 3g). The sprouting defect was also not due to reduced EC viability or increased oxidative stress, or to changes in energy charge, glutathione or NADPH levels, glycolysis, glucose or glutamine oxidation, or oxygen consumption (Extended Data Fig. 2d–m).

$GLUL^{KD}$ ECs showed impaired migration in scratch-wound and Boyden chamber assays, even upon treatment with mitomycin C, an effect that was rescued by re-introducing the shRNA-resistant *GLUL* ($rGLUL^{OE}$) (Fig. 3h, i). Furthermore, sparsely seeded $GLUL^{KD}$ ECs had a reduced velocity of random movement (Fig. 3j, Supplementary Videos 1, 2) and a decreased lamellipodial area (Fig. 3k–m). Comparable results were obtained with a second non-overlapping shRNA and a *GLUL*-specific small interfering RNA (siRNA) (Extended Data Figs. 2a, 3a–e).

The migration defects suggested that the remodelling of the actin cytoskeleton, which is necessary for cellular motility, was perturbed in $GLUL^{KD}$ ECs. Notably, we detected an increase in F-actin levels in $GLUL^{KD}$ ECs (Fig. 3n). A role for GS in cytoskeletal remodelling was further suggested by analysing the repolymerization of the actin cytoskeleton upon disruption with the F-actin polymerization inhibitor latrunculin B and subsequent wash-out. Latrunculin B perturbed the normal morphology of control and $GLUL^{KD}$ ECs (Fig. 3o–r). After wash-out, when control cells had rebuilt a normal actin cytoskeleton, $GLUL^{KD}$ ECs still had higher F-actin levels, which mainly originated from increased numbers of stress-fibre bundles (Fig. 3s–u). $GLUL^{KD}$ ECs did not display altered α-tubulin levels (Fig. 3v, Extended Data Fig. 4a–h).

The increase in F-actin levels was also present in ECs that were freshly isolated from MSO-treated mice (Extended Data Fig. 4i–k), and in confluent $GLUL^{KD}$ ECs aligning a scratch wound in vitro (Extended Data Fig. 4l–n). Confluent monolayer $GLUL^{KD}$ ECs displayed compromised junctional integrity (Extended Data Fig. 4o–v). Functionally, this corresponded to a decrease in the transendothelial electrical resistance of $GLUL^{KD}$ ECs in vitro (Extended Data Fig. 4w) and increased leakiness of inflamed (but not healthy) vessels in vivo (Extended Data Fig. 4x–z).

Glutamine production by endothelial GS

To explore whether the migration defect was attributable to reduced de novo glutamine synthesis, we measured the glutamine-synthesizing activity of GS by supplementing ECs with $^{15}NH_4Cl$ (Extended Data Fig. 5a). At a physiological concentration of 0.6 mM glutamine or higher, the glutamine-producing activity of GS was negligible, at approximately the level observed in ECs treated with MSO. It slightly increased only upon glutamine withdrawal, presumably to compensate for the lack of available glutamine (Fig. 4a). Similar results were obtained in a medium containing dialysed serum (Extended Data Fig. 5b). For further details see Supplementary Discussion 1 and Extended Data Fig. 5c–n.

To determine whether the phenotype of $GLUL^{KD}$ ECs relied on the catalytic site of GS, we used MSO—which is an irreversible inhibitor of GS by competing with glutamate for binding at the catalytic site—at previously reported concentrations[10]. MSO reduced EC spheroid sprouting, impaired EC migration in scratch-wound assays under treatment with mitomycin C, and decreased lamellipodial area, while increasing F-actin levels after latrunculin B wash-out but not affecting EC proliferation (Extended Data Fig. 5o–t). Even though other (off-target) effects of pharmacological GS inhibition cannot be

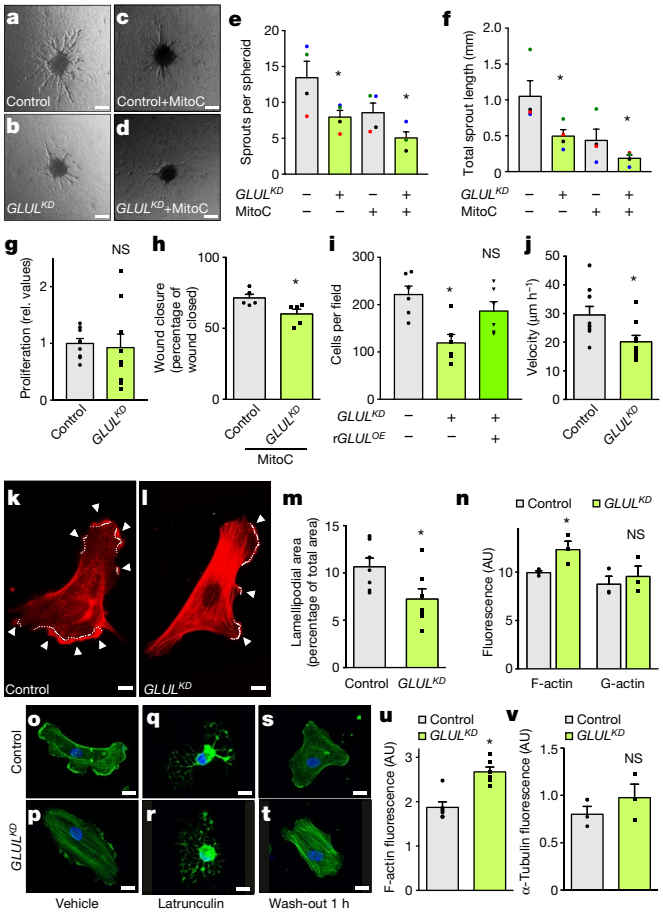


Fig. 3 | Loss of *GLUL* impairs EC migration through perturbed actin dynamics. a–f, Control (**a**, **c**) and $GLUL^{KD}$ (**b**, **d**) EC spheroids without (**a**, **b**) and with (**c**, **d**) mitomycin C (MitoC) treatment, and number of sprouts per spheroid (**e**) and total sprout length (**f**). **g**, Proliferation of control and $GLUL^{KD}$ ECs, as measured by [^{3}H]thymidine incorporation. **h**, Wound closure upon treatment of control and $GLUL^{KD}$ ECs with mitomycin C. **i**, Boyden chamber migration for control, $GLUL^{KD}$ and $GLUL^{KD}$ + r$GLUL^{OE}$ (overexpression of a shRNA-resistant *GLUL* mutant) ECs, all under mitomycin C treatment. **j**, Velocity of sparsely seeded control and $GLUL^{KD}$ ECs. **k–m**, Phalloidin (F-actin) staining of control (**k**) and $GLUL^{KD}$ (**l**) ECs (arrows and white dotted lines indicate lamellipodia) and quantification of lamellipodial area (**m**). **n–p**, F-actin and G-actin levels in phalloidin (F-actin)–DNase I (G-actin) double-stained control and $GLUL^{KD}$ ECs (**n**), and representative images of phalloidin-stained control (**o**) and $GLUL^{KD}$ (**p**) ECs. **q–u**, Phalloidin staining of latrunculin B-treated control (**q**, **s**) and $GLUL^{KD}$ (**r**, **t**) ECs at time point 0 (**q**, **r**) and at 1 h after latrunculin wash-out (**s**, **t**) and quantification of F-actin levels after wash-out (**u**). **v**, α-Tubulin levels in $GLUL^{KD}$ and control ECs. All data are mean ± s.e.m., with individual data points shown for **e–j**, **m**, **n**, **u** and **v**; values of *n* (independent experiments) are: 4 (**e**, **f**), 9 (**g**, **j**), 5 (**h**), 6 (**i**, **u**), 7 (**m**) and 3 (**n**, **v**). NS, $P > 0.05$, $^*P < 0.05$ according to mixed-models R statistics (**e**, **f**), Student's *t*-test (**g**, **h**, **j**, **m**, **n**, **u**, **v**) or one-way ANOVA with Dunnett's multiple comparison versus control (**i**). Exact *P* values are as follows: control versus $GLUL^{KD}$ ± mitoC: <0.0001 (**e**, **f**); 0.7729 (**g**); 0.0283 (**h**); control versus $GLUL^{KD}$: 0.0093; control versus $GLUL^{KD}$ + r$GLUL^{OE}$: 0.5981 (**i**); 0.0234 (**j**); 0.0352 (**m**); F-actin: 0.0467; G-actin: 0.584 (**n**); 0.0007 (**u**); 0.3491 (**v**). AU, arbitrary units × 10^3. Scale bars: 100 µm (**a–d**), 10 µm (**k**, **l**) and 20 µm (**o–t**).

formally excluded, MSO phenocopied the *GLUL* knockdown, which suggests that the catalytic site of GS is indispensable in the control of EC cytoskeletal homeostasis.

GS inhibition affects RHOJ activity

Because small GTPases and their effectors control F-actin levels and motility[11], we explored whether Rho GTPases were downstream targets of GS. We focused on RHOJ, because it is mainly expressed in ECs[12] and because the blocking of endothelial RHOJ has been proposed as a novel anti-angiogenesis approach[13]. It is noteworthy that $RHOJ^{KD}$ ECs fully phenocopied $GLUL^{KD}$ ECs in terms of decreased mobility and barrier function (data not shown).

Because RHOJ localizes to plasma and organelle membranes to become activated[14], and is almost exclusively detected in the membrane fraction[15], we investigated whether its membrane localization and activity were regulated by GS levels. Immunoblotting revealed that RHOJ was detectable only in the membrane fraction of ECs (consistent with previous findings[15]), and that $GLUL^{KD}$ ECs had decreased amounts of RHOJ in the membrane fraction (without concomitant increase in the cytosolic fraction, possibly because of proteasomal degradation[16]) as well as decreased levels of active RHOJ (Fig. 4b, c). *GLUL* knockdown did not clearly affect *RHOJ* transcript levels (relative mRNA levels: 0.99 ± 0.03 in control compared with 0.85 ± 0.05 in $GLUL^{KD}$; $n = 3$, $P = 0.0282$).

We also explored whether *GLUL* knockdown affected other Rho GTPases in ECs. We focused on the RHOA/B/C–Rho kinase (ROCK)–myosin light chain (MLC) axis, because silencing of endothelial RHOJ increases signalling of this pathway and induces aberrant F-actin stress-fibre formation through an as-yet-undefined mechanism[13,17] (Fig. 4d). Standard glutathione *S*-transferase–Rhotekin pull-down assays showed that *GLUL* knockdown increased the activity of RHOA and RHOC, but not of RHOB (Fig. 4e–g). Of note, *GLUL* knockdown, much like other stimuli, increased total RHOB levels. We confirmed the increase in RHOA activity at the individual-cell level with a DORA–RHOA–FRET (DORA, dimerization optimized reporter for activation; FRET, fluorescence resonance energy transfer) biosensor (Fig. 4h, Extended Data Fig. 6a). We observed that the abnormally elevated RHOA activity in retracting lamellipodia in $GLUL^{KD}$ ECs evoked more numerous, but smaller and more short-lived, lamellipodia (Fig. 4i), which could contribute to the motility impairment. As previously suggested[18], increased RHOA activity in lamellipodia leads to local actomyosin contraction through ROCK and phosphorylated MLC (pMLC), thereby prematurely retracting the lamellipodium. Combining both *GLUL* and *RHOJ* knockdown did not further increase RHOA activity (data not shown); this confirms that RHOJ silencing by itself increased RHOA activity, and suggests that GS primarily acts via RHOJ to control RHOA signalling.

Downstream of Rho GTPases, $GLUL^{KD}$ and MSO-treated ECs had increased ROCK1 and ROCK2 protein levels (Fig. 4j) and enhanced ROCK activity, as determined by pMLC protein levels, which were similarly induced in both $GLUL^{KD}$ and $RHOJ^{KD}$ ECs (Fig. 4k, Extended Data Fig. 6b–n). In agreement with this, ROCK inhibitors (Y27632, fasudil hydrochloride and H1152 dihydrochloride (data not shown)) rescued the phenotype associated with $GLUL^{KD}$ (Fig. 4l–o, Extended Data Fig. 6o–w), whereas myosin light-chain kinase inhibitors (ML7; peptide 18) did not (Extended Data Fig. 6x–aa). This suggests that MLC phosphorylation through ROCK rather than through myosin light-chain kinase is more important in mediating the phenotype associated with $GLUL^{KD}$ in ECs. Therefore, *GLUL* knockdown reduces the membrane localization and activity of RHOJ, while activating RHOA, RHOC and ROCK.

We next explored which of these Rho GTPases interact with GS, assuming that such an interaction might facilitate and/or be necessary for their activation; we remained mindful, however, that RHOJ can negatively regulate the activity of the RHOA–ROCK–MLC axis[13,17], and hence the loss of a primary interaction of GS with RHOJ could indirectly explain the increased levels and/or activity of RHOA, ROCK and pMLC upon *GLUL* knockdown. First, co-immunoprecipitation assays revealed an interaction between endogenous RHOJ and GS (Fig. 5a); such an interaction was not observed for RHOA or for RHOC, which is the most abundant Rho GTPase in ECs (Extended Data Fig. 7a). Second, deletion of the first 20 N-terminal amino acids of RHOJ (ΔN20-RHOJ), which mediate its plasma-membrane localization[19], reduced the interaction with GS (Extended Data Fig. 7b). Third,

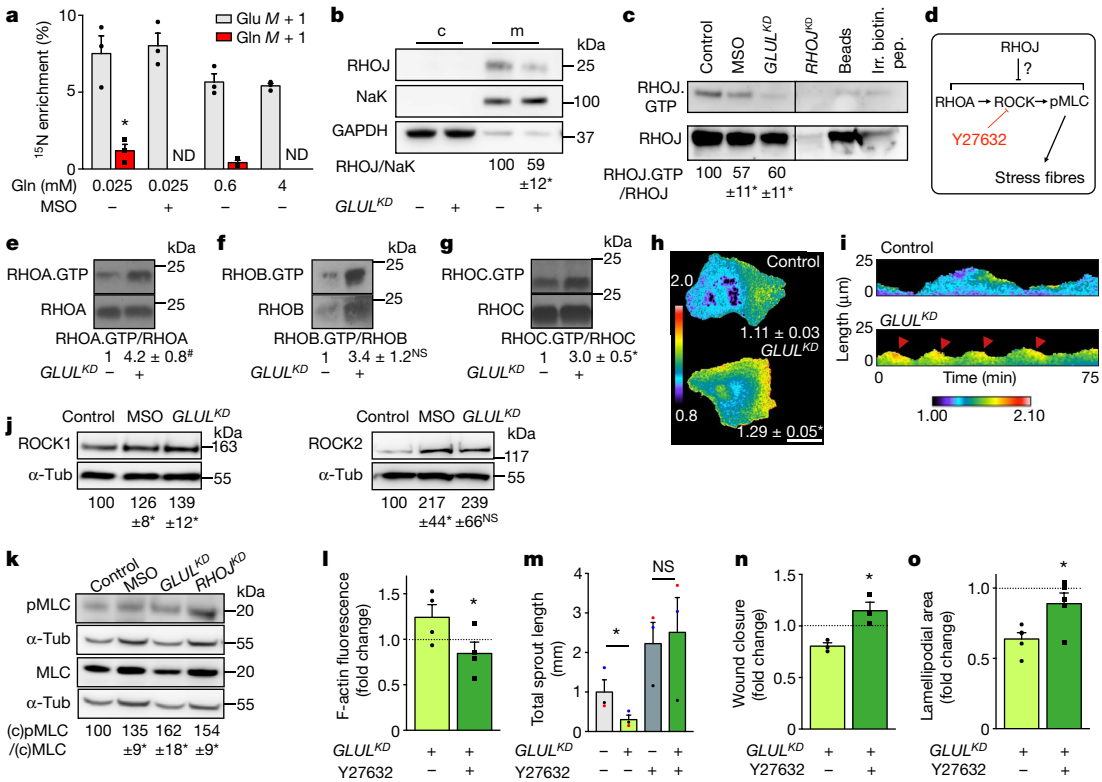


Fig. 4 | Endothelial GS regulates Rho GTPase activity. a, Effect of glutamine and MSO on glutamine-producing activity, measured as the percentage enrichment of $M+1$ (singly ^{15}N-labelled) glutamine (Gln) and glutamate (Glu), 30 min after the addition of $^{15}NH_4^+$. **b**, RHOJ, NaK ATPase (membrane marker) and GAPDH (cytosol marker) immunoblots in cytosolic (c) and membrane (m) fractions with quantification. **c**, Immunoblot for active and total RHOJ with quantification; *RHOJ*^{KD}, beads only (Beads) and irrelevant biotinylated peptide (Irr. biotin. pep.) are negative controls. **d**, The pivotal yet incompletely understood role of RHOJ in EC migration and stress-fibre formation. **e**–**g**, Immunoblots for pull-down RHOA (**e**), RHOB (**f**) and RHOC (**g**) activity assays with quantification. **h**, Control and *GLUL*^{KD} ECs expressing the DORA–RHOA biosensor, with quantification of whole-cell FRET start ratio (mean ± s.e.m.; control, $n=12$ cells; *GLUL*^{KD}, $n=9$). Look-up table (colour bar) denotes relative RHOA activities (blue, low; red, high). **i**, Kymograph of DORA–RHOA biosensor-expressing ECs, showing abnormally short-lived lamellipodia and increased RHOA activity in retracting lamellipodia of *GLUL*^{KD} ECs (red arrowheads) (representative of 13 control and *GLUL*^{KD} cells). **j**, Immunoblots of ROCK1 (left), ROCK2 (right) and α-tubulin (α-Tub), with quantification. **k**, pMLC, total MLC and α-tubulin immunoblots (for quantification, see Methods). In this panel, parenthetical '(c)' denotes 'corrected for corresponding loading control'. **l**, F-actin levels after latrunculin B wash-out in ECs treated with the ROCK inhibitor Y27632. **m**–**o**, Effect of Y27632 on spheroid-sprouting defect (**m**), migration defect (**n**), and lamellipodial area (**o**). Values in **l**, **n**, **o** are relative to untreated non-silenced control (dotted line). Scale bar, 25 µm (**h**). All data are mean ± s.e.m., individual data points are shown for **a** and **l**–**o**; values of n (independent experiments) are: 3 (**a**, **e**, **f**, **m**, **n**), 4 (**c** (MSO), **h**, **k**, **l**), 5 (**o**), 7 (**j**), 8 (**c** (*GLUL*^{KD}), **g**), 13 (**b**). NS, $P>0.05$, $^{\#}P=0.05$, $^{*}P<0.05$; one-way ANOVA with Dunnett's multiple comparisons versus 4 mM (**a**), one-sample t-test (**b**, **c**, **e**, **f**, **g**, **j**, **k**), Student's t-test (**h**, **n**, **o**), paired Student's t-test (**l**) or mixed-models R statistics (**m**). Exact P values are as follows: (Glu) 0.6 mM versus 4 mM: 0.9903; 0.025 mM + MSO versus 4 mM: 0.0968; 0.025 mM versus 4 mM: 0.1943; (Gln) 0.6 mM versus 4 mM: 0.4518; 0.025 mM + MSO versus 4 mM: 0.9999; 0.025 mM versus 4 mM: 0.0143 (**a**); 0.0072 (**b**); MSO: 0.0323; *GLUL*^{KD}: 0.0095 (**c**); 0.053 (**e**); 0.1790 (**f**); 0.0035 (**g**); 0.0055 (**h**); ROCK1 MSO: 0.0169; ROCK1 *GLUL*^{KD}: 0.0138; ROCK2 MSO: 0.0381; ROCK2 *GLUL*^{KD}: 0.0802; (**k**) MSO: 0.0283; *GLUL*^{KD}: 0.0431; *RHOJ*^{KD}: 0.0091 (**j**); 0.0431 (**l**); *GLUL*^{KD} versus control: <0.0001; *GLUL*^{KD} + Y27632 versus control + Y27632: 0.5211 (**m**); 0.0181 (**n**); 0.0210 (**o**). For gel source images, see Supplementary Fig. 1.

immunoblotting showed that only RHOJ, and not RHOA or RHOC, was predominantly membrane-localized (Extended Data Fig. 7c). Fourth, we confirmed the GS–RHOJ interaction with a bimolecular fluorescence complementation approach (Extended Data Fig. 7d, e). On the basis of the above data, we focused on RHOJ as the most likely interaction partner of GS.

To interact with membrane-localized (active) RHOJ, GS should be similarly located; indeed, cell fractionation studies revealed that a fraction of GS was membrane-localized (Fig. 5b). Further evidence is derived from single-particle tracking data, acquired by photoactivated localization microscopy imaging (SPT-PALM) combined with total internal reflection fluorescence microscopy (TIRF). We traced the movement of single GS proteins that were tagged with the photoswitchable fluorescent protein mEOS (GS–mEOS). Single GS–mEOS particles had a lower diffusion coefficient in the TIRF region (comprising the plasma membrane and the immediately adjacent cytoplasm) than did free mEOS, which is indicative of an association of GS with membrane structures (Fig. 5c, Extended Data Fig. 7f).

Palmitoylation of GS and RHOJ

Membrane localization often requires post-translational palmitoylation. We therefore proposed that GS could be palmitoylated to enable plasma membrane localization and interaction with RHOJ. We performed click chemistry with biotin azide (Extended Data Fig. 7g) on lysates from HEK293 cells that overexpressed GS and had been treated with the clickable palmitoylation probes 16C-BYA or 16C-YA. Subsequent streptavidin pull-down showed that both probes labelled GS, which was a clear indication that GS had been palmitoylated. The labelling was reduced by MSO, which is consistent with the presumed dependency of the phenotype on the catalytic site of the enzyme (Fig. 5d).

The palmitoylation of GS has previously been reported anecdotally, however in-depth molecular and functional characterization was not carried out[20]. To determine whether GS undergoes autopalmitoylation, we incubated purified GS[21] with palmitoyl-alkyne coenzyme A (palmitoyl-alkyne CoA; a substrate for palmitoylation) in a cell-free system without any other proteins present, to demonstrate a direct effect. Click

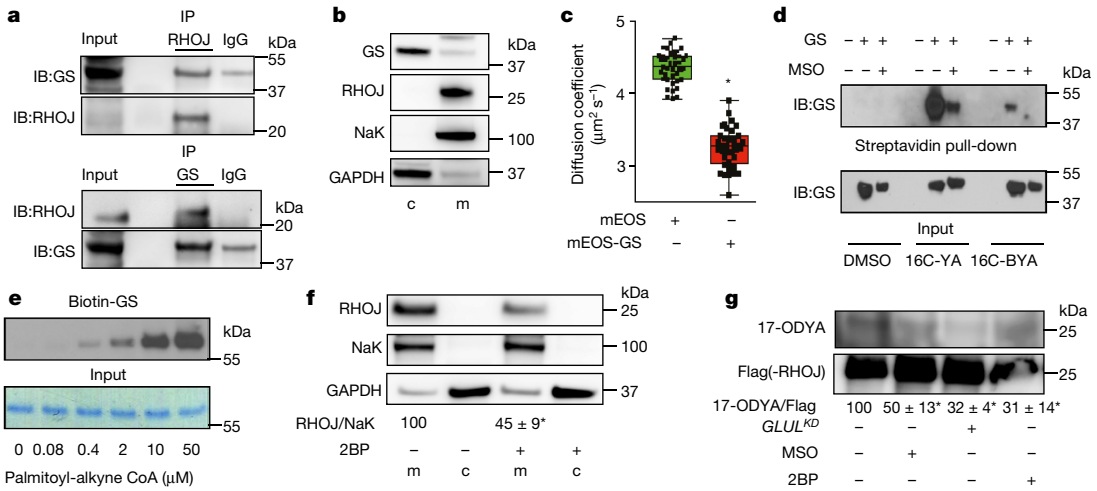


Fig. 5 | GS (auto)-palmitoylation. a, Co-immunoprecipitation of endogenous RHOJ and GS in ECs. Top, immunoprecipitation of RHOJ; bottom, immunoprecipitation of GS. **b**, Immunoblot for GS and RHOJ in cytosolic (c) and membrane (m) fractions in ECs with NaK and GAPDH as fraction markers. **c**, Diffusion coefficient of single photoswitchable fluorescent protein mEOS and mEOS-fused GS (mEOS–GS) particles in the plasma-membrane region of ECs acquired by SPT-PALM under TIRF illumination ($n = 41$ cells expressing mEOS and 37 expressing mEOS–GS). **d**, GS immunoblotting after streptavidin pull-down of biotin-azide-clicked lysates from HEK-293T cells for the indicated palmitoylation probes. Input shows levels of GS overexpression. **e**, Effect of the concentration of palmitoyl-alkyne CoA on the autopalmitoylation of purified GS; biotin-azide clicking and HRP-streptavidin blotting are shown, with the input shown on Coomassie-stained gel. **f**, Immunoblotting for RHOJ, NaK and GAPDH in membrane (m) and cytosolic (c) fractions of control- and 2-bromopalmitate (2BP)-treated ECs. **g**, Palmitoylation of RHOJ in *GLUL*^{KD}, MSO- and 2BP-treated ECs. In-gel fluorescence for tetramethylrhodamine (TAMRA)-azide 17-ODYA (palmitoylation probe)-clicked Flag–RHOJ is shown (Flag as loading control). 2BP is a pan-palmitoylation inhibitor. All data are mean ± s.e.m., except box and whisker (running from minimal to maximal values) plots in (**c**), for which individual data points are shown; values of n (independent experiments) are: 2 (**e**), 3 (**a**, **b**, **c**, **d**, **f**), 4 (**g**). NS, $P > 0.05$, *$P < 0.05$; Student's t-test (**c**); one-sample t-test (**f**, **g**). Exact P values are as follows: <0.0001 (**c**); 0.0264 (**f**); MSO: 0.0317; *GLUL*^{KD}: 0.0003; 2BP: 0.0163 (**g**). For gel source images, see Supplementary Fig. 1.

chemistry revealed that increasing the dose of palmitoyl-alkyne CoA resulted in increased autopalmitoylation of GS (Fig. 5e). Importantly, autopalmitoylation of GS was achieved with physiological concentrations of palmitoyl-alkyne CoA (1–10 μM) at neutral pH, which suggests physiologically relevant autopalmitoylation; this was subsequently confirmed with two alternative methods (Supplementary Discussion 2 and Extended Data Fig. 7h–j).

Palmitoylation of target proteins by palmitoyl-acyl transferases is a two-step reaction, requiring first autopalmitoylation of the palmitoyl-acyl transferase, followed by transfer of the palmitoyl group to the target protein. We considered that GS could have a similar activity profile (Supplementary Discussion 3) and explored whether it was involved in the palmitoylation of RHOJ. Even though the cysteines at positions 3 (C3) and 11 (C11) of RHOJ were predicted by in silico methods to be high-fidelity palmitoylation sites (screened with SwissPalm[22], data not shown), the palmitoylation of RHOJ has been poorly documented, with the exception of a few studies[23,24]. Notably, the membrane-localization and activity of RHOJ were reduced by treatment of ECs with the pan-palmitoylation inhibitor 2-bromopalmitate and by introducing point mutations in C3 and C11 (Fig. 5f, Extended Data Fig. 7k–t), providing initial evidence that RHOJ can be palmitoylated in ECs. Using the palmitoylation probe 17-ODYA (Fig. 5g) or an acyl-resin-assisted capture (Extended Data Fig. 7u), we found a reduction in the levels of palmitoylated RHOJ upon blocking GS, which is consistent with a model whereby GS sustains palmitoylation of RHOJ.

Discussion

We have found that GS is active in the regulation of EC motility. This activity is presumably independent of glutamine synthesis, although we cannot formally exclude a possible contribution of minimal levels of glutamine production by GS to the observed phenotype. GS was found to regulate RHOJ signalling in cell motility, as shown by several forms of evidence. First, a fraction of GS is present in EC membranes, which is where active RHOJ resides. Second, GS interacts with RHOJ in ECs in co-immunoprecipitation experiments, although this interaction could be direct or indirect. Third, *GLUL* knockdown reduces the palmitoylation of RHOJ, its membrane localization and its activity in ECs. Therefore, because RHOJ promotes EC motility[13,17], the impaired migration of *GLUL*^{KD} ECs could be attributed to a reduction in RHOJ activity. RHOJ probably also indirectly contributes to promoting EC motility by controlling the activity of the RHOA–ROCK–MLC signalling pathway, which is known to regulate EC motility by affecting stress-fibre formation[13,17] (Extended Data Fig. 7v, Supplementary Discussion 4).

Because purified GS seems to be capable of autopalmitoylation—a feature of palmitoyl-acyl transferase enzymes—and because *GLUL* silencing decreases the palmitoylation of RHOJ, our data support a model whereby GS first palmitoylates itself and then transfers the palmitoyl group to RHOJ, although we cannot formally exclude the possibility that transfer of the palmitoyl group from GS to RHOJ occurs via additional partners or even non-enzymatically. A possible model for GS palmitoylation is described in Supplementary Discussion 5, Extended Data Fig. 8 and Extended Data Table 2. In addition, whether GS interacts exclusively with RHOJ or whether it can interact with other proteins (for example, other palmitoylated Rho GTPases such as RAC1, CDC42, RHOU or RHOV) to mediate this effect on EC motility is yet to be clarified. In any case, RHOJ seems to be a critical target of GS, given that its silencing completely phenocopies GS inhibition in ECs.

Finally, GS is critical for EC motility and migration, which contributes to the formation of new vessels in development and disease. By contrast, ECs do not migrate when they are quiescent in healthy adults, which explains why GS inhibition has no observable effects on the vasculature in healthy adult mice. This renders GS an attractive disease-restricted target for the therapeutic inhibition of pathological angiogenesis. Furthermore, the pharmacological GS blocker MSO reduced pathological angiogenesis in blinding eye and psoriatic skin disease (Fig. 2), which warrants further exploration of GS targeting in anti-angiogenesis.

Online content

Any Methods, including any statements of data availability and Nature Research reporting summaries, along with any additional references and Source Data files, are available in the online version of the paper at https://doi.org/10.1038/s41586-018-0466-7.

Received: 8 January 2016; Accepted: 18 July 2018;
Published online: 29 August 2018

1. De Bock, K. et al. Role of PFKFB3-driven glycolysis in vessel sprouting. *Cell* **154**, 651–663 (2013).
2. Schoors, S. et al. Fatty acid carbon is essential for dNTP synthesis in endothelial cells. *Nature* **520**, 192–197 (2015).
3. Schoors, S. et al. Partial and transient reduction of glycolysis by PFKFB3 blockade reduces pathological angiogenesis. *Cell Metab.* **19**, 37–48 (2014).
4. Huang, H. et al. Role of glutamine and interlinked asparagine metabolism in vessel formation. *EMBO J.* **36**, 2334–2352 (2017).
5. Abcouwer, S. F., Lukascewicz, G. C., Ryan, U. S. & Souba, W. W. Molecular regulation of lung endothelial glutamine synthetase expression. *Surgery* **118**, 325–335 (1995).
6. He, Y., Hakvoort, T. B., Vermeulen, J. L., Lamers, W. H. & Van Roon, M. A. Glutamine synthetase is essential in early mouse embryogenesis. *Dev. Dyn.* **236**, 1865–1875 (2007).
7. Häberle, J. et al. Congenital glutamine deficiency with glutamine synthetase mutations. *N. Engl. J. Med.* **353**, 1926–1933 (2005).
8. Kung, H. N., Marks, J. R. & Chi, J. T. Glutamine synthetase is a genetic determinant of cell type-specific glutamine independence in breast epithelia. *PLoS Genet.* **7**, e1002229 (2011).
9. Schoors, S. et al. Incomplete and transitory decrease of glycolysis: a new paradigm for anti-angiogenic therapy? *Cell Cycle* **13**, 16–22 (2014).
10. Dadsetan, S. et al. Brain alanine formation as an ammonia-scavenging pathway during hyperammonemia: effects of glutamine synthetase inhibition in rats and astrocyte-neuron co-cultures. *J. Cereb. Blood Flow Metab.* **33**, 1235–1241 (2013).
11. Ridley, A. J. Rho GTPase signalling in cell migration. *Curr. Opin. Cell Biol.* **36**, 103–112 (2015).
12. Yuan, L. et al. RHOJ is an endothelial cell-restricted Rho GTPase that mediates vascular morphogenesis and is regulated by the transcription factor ERG. *Blood* **118**, 1145–1153 (2011).
13. Kim, C. et al. Vascular RHOJ is an effective and selective target for tumor angiogenesis and vascular disruption. *Cancer Cell* **25**, 102–117 (2014).
14. Leszczynska, K., Kaur, S., Wilson, E., Bicknell, R. & Heath, V. L. The role of RHOJ in endothelial cell biology and angiogenesis. *Biochem. Soc. Trans.* **39**, 1606–1611 (2011).
15. de Toledo, M. et al. The GTP/GDP cycling of Rho GTPase TCL is an essential regulator of the early endocytic pathway. *Mol. Biol. Cell* **14**, 4846–4856 (2003).
16. Nethe, M. & Hordijk, P. L. The role of ubiquitylation and degradation in RhoGTPase signalling. *J. Cell Sci.* **123**, 4011–4018 (2010).
17. Kaur, S. et al. RHOJ/TCL regulates endothelial motility and tube formation and modulates actomyosin contractility and focal adhesion numbers. *Arterioscler. Thromb. Vasc. Biol.* **31**, 657–664 (2011).
18. Heasman, S. J. & Ridley, A. J. Multiple roles for RHOA during T cell transendothelial migration. *Small GTPases* **1**, 174–179 (2010).
19. Ackermann, K. L., Florke, R. R., Reyes, S. S., Tader, B. R. & Hamann, M. J. TCL/RHOJ plasma membrane localization and nucleotide exchange is coordinately regulated by amino acids within the N terminus and a distal loop region. *J. Biol. Chem.* **291**, 23604–23617 (2016).
20. Wan, J. et al. Tracking brain palmitoylation change: predominance of glial change in a mouse model of Huntington's disease. *Chem. Biol.* **20**, 1421–1434 (2013).
21. Levine, R. L., Oliver, C. N., Fulks, R. M. & Stadtman, E. R. Turnover of bacterial glutamine synthetase: oxidative inactivation precedes proteolysis. *Proc. Natl Acad. Sci. USA* **78**, 2120–2124 (1981).
22. Blanc, M. et al. SwissPalm: Protein Palmitoylation database. *F1000Res.* **4**, 261 (2015).
23. Roberts, P. J. et al. Rho family GTPase modification and dependence on CAAX motif-signaled posttranslational modification. *J. Biol. Chem.* **283**, 25150–25163 (2008).
24. Wei, X., Song, H. & Semenkovich, C. F. Insulin-regulated protein palmitoylation impacts endothelial cell function. *Arterioscler. Thromb. Vasc. Biol.* **34**, 346–354 (2014).

Acknowledgements We acknowledge R. Levine for supplying purified bacterial GS; L. Van Den Bosch and W. Scheveneels for providing primary mouse astrocytes; S. Trenson, I. Crèvecoeur, S. Noppen, L. Van Berckelaer and R. Van Berwaer for technical assistance; S.-M. Fendt, D. Verdegem and C. Ulens for discussions and suggestions; W. Vermaelen, A. Acosta Sanchez, A. Brajic, A. Sobrino and M. Cockx for experimental assistance; L.-C. Conradi and A. Pircher for supplying materials; and E. Wauters, A. Wolthuis and J. Jaekers for providing tissues for EC isolations. HecBioSim and PRACE are acknowledged for allocation of computer time. J.G., A.R.C., C.D., C.L., J.K., F.M.-R., S.R. and S.Va. are supported by the FWO; A.Z. by LE&RN/FDRS; B.C. by IWT; U.B. by a Marie Curie-IEF Fellowship; H.H. by an EMBO Long-Term Fellowship; J.D.v.B. by a LSBR fellowship; R.Cu. by a British Heart Foundation Intermediate Clinical Fellowship; and X.W. by the American Cancer Society RSG, NIH/NCI and NIH/NIDDK. F.C., G.S. and F.L.G. are supported by the EPSRC; X.L. by the State Key Laboratory of Ophthalmology, Zhongshan Ophthalmic Center at Sun Yat-Sen University, and the National Natural Science Foundation of China (81330021 and 81670855). P.C. is supported by a Federal Government Belgium grant, long-term structural Methusalem funding by the Flemish Government, a Concerted Research Activities Belgium grant, grants from the FWO, Foundation against Cancer and ERC Advanced Research Grant. G.E. and M.Dew. received a Foundation against Cancer grant and G.E. received a FWO 'Krediet aan navorsers'.

Author contributions P.C. conceived the concept of the study and provided supervision. G.E., P.C., X.L., M.Dew. and L.S. contributed to the execution, support and analysis of experiments, and/or provided advice. G.E., C.D., A.R.C., J.G. and P.C. were involved in the experimental design. G.E., C.D., A.R.C., J.G., U.B., A.Z., H.H., S.Va., J.K., C.L., F.M.-R., B.C., L.R., S.Vi., K.B., S.W., J.S., L.S., S.L., R.Ch., R.Cu. and M.Dew. carried out molecular biology and in vivo experiments. B.G. performed mass spectrometry. J.v.R. and J.D.v.B. carried out RHO activity assays. M.DeR., G.J. and X.W. undertook palmitoylation experiments. G.S., F.C. and F.L.G. performed molecular dynamics simulations. S.R. and J.Ho. performed bimolecular fluorescence complementation and SPT experiments. G.E., C.D., A.R.C., J.G., R.Cu., U.B., C.L., S.R., L.T., B.C., M.Dew., J.Ho., S.L., B.G., F.L.G., J.D.v.B., X.W. and P.C. interpreted the data. W.H.L., Y.W. and J.Ha. provided necessary materials. G.E. and P.C. drafted the manuscript. All authors agreed on the final version of the manuscript.

Competing interests The authors declare no competing interests.

Additional information
Extended data is available for this paper at https://doi.org/10.1038/s41586-018-0466-7.
Supplementary information is available for this paper at https://doi.org/10.1038/s41586-018-0466-7.
Reprints and permissions information is available at http://www.nature.com/reprints.
Correspondence and requests for materials should be addressed to G.E. or X.L. or P.C.
Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

METHODS

Chemicals and reagents. The GS inhibitor L-methionine sulfoximine (MSO), mitomycin C, latrunculin B, oligomycin, antimycin A, carbonyl cyanide-4-(trifluoromethoxy)phenylhydrazone (FCCP), 2-bromohexadecanoic acid (2-bromopalmitic acid, 2BP), tamoxifen, palmitoyl-CoA agarose and α-ketoglutarate dehydrogenase were from Sigma-Aldrich. 17-Octadecynoic acid (17-ODYA) and biotin-azide were purchased from Cayman Chemical. The use and/or synthesis of the other palmitoylation probes 15-hexadecynoic acid (16C-YA; a palmitate-based probe that binds a broader spectrum of proteins than does 16C-BYA (below), including both palmitoyl-acyl transferases and palmitoyl-acyl transferase target proteins) and 2-bromooctadec-15-yonic acid (16C-BYA; a 2-bromopalmitate-based activity-based probe that labels but also inhibits palmitoyl-acyl transferase enzymes) has previously been described[25]. The ROCK kinase inhibitor Y27632 ((*R*)-(+)-*trans*-4-(1-aminoethyl)-*N*-(4-pyridyl)cyclohexanecarboxamide) was from BioVision, fasudil hydrochloride and H1152 dihydrochloride were from Tocris. The myosin light-chain kinase inhibitors ML7-hydrochloride and peptide 18 were from Tocris. Collagen type 1 (rat tail) was obtained from Merck Millipore. [5-^{3}H]glucose, [^{3}H]thymidine and [U-^{14}C]glutamine were from Perkin Elmer; [6-^{14}C]D-glucose was from ARC. [U-^{13}C]glucose, [U-^{13}C]glutamine, [U-^{13}C]glutamate and $^{15}NH_4Cl$ were purchased from Cambridge Isotope Laboratories. The following primary antibodies or dyes were used (dilutions for staining (ST), immunoblotting (IB), immunofluorescence (IF) and immunoprecipitation (IP) are given between brackets): *Griffonia simplicifolia* (GS)-IB_4-Alexa 488 (ST 1:200), isolectin GS-IB_4-Alexa 568 (ST 1:200), isolectin GS-IB_4-Alexa 647 (ST 1:200), phalloidin-Alexa 488 (ST 1:100), DNase I-Alexa 594 (ST 1:200) (Molecular Probes), anti-collagen IV (2150–1470) (IF 1:400) (BioRad), anti-NG2 chondroitin sulfate proteoglycan (AB5320) (IF 1:200) (Millipore), anti-Flag (clone M2) (IB 1:1,000; IP 5 μg ml^{-1}), anti-GS (clone 2B12) (IB 1:1,000; IP 2–5 μg ml^{-1}), anti-RHOJ (clone 1E4) (IB 1:1,000; IP 2–5 μg ml^{-1}), anti-ROCK1 (HPA007567) (IB 1:1,000), anti-α-tubulin (T6199) (IB 1:1,000) (Sigma-Aldrich), anti-β-actin (13E5) (IB 1:1,000), anti-phospho-myosin light-chain 2 (IB 1:1,000; IF 1:300) and anti-myosin light-chain 2 (IB 1:1,000) (9776), anti-Na,K-ATPase (NaK) (3010) (IB 1:1,000), anti-RHOA (67B9) (IB 1:1,000) and anti-RHOC (D40E4) (IB 1:1,000) (Cell Signaling Technology), anti-CD105/endoglin (AF1320) (IF 1:50), anti-VE-cadherin (AF1002) (IF 1:50) (R&D Systems), anti-ROCK2 (A300-047A-T) (IB 1:500) (Imtec Diagnostics), anti-CD31 (MEC13.3) (IF 1:200), anti-CD34-biotin (553732) (IF 1:25) (BD Biosciences), anti-RHOB (sc-180) (IB 1:1,000). Secondary Alexa-405, -488, -568 or -647-conjugated antibodies (1:500) were from Molecular Probes; other secondary antibodies and immunoglobulin-γ (IgG) controls were from Dako. The Click-iT EdU Alexa Fluor 555 Imaging Kit was from Invitrogen. Purified bacterial GS was a gift from R. Levine (Bethesda).

Cell culture. *HUVECs and human umbilical artery ECs.* The cells were obtained under protocol S57123 (Commission Medical Ethics of UZ/KU Leuven) after written consent of the donors, were isolated as previously described[1,2] and were routinely cultured in M199 medium (Invitrogen) containing 20% FBS, 0.6 mM L-glutamine, heparin (10 U ml^{-1}; Sigma), penicillin (100 U ml^{-1}), streptomycin (100 μg ml^{-1}) and endothelial cell growth factor supplements (ECGS; 30 mg l^{-1}; Sigma). Cells were only used between passages 1 and 4 and all experiments were performed in HUVECs from at least three different donors unless stated otherwise. Except when stated otherwise, the use of the abbreviation EC in the text refers to HUVEC.

Isolation of endothelial cells from human lung, liver or colon mucosa. Lung, liver or colon mucosa specimens were obtained under protocol S57123 (Commission Medical Ethics of UZ/KU Leuven) and were washed several times with phosphate buffer solution (PBS) and minced with scissors before enzymatic digestion for 45 min at 37 °C with collagenase/dispase/DNase solution (Gibco, Life Technologies). The resulting suspension was passed through a 100-μm nylon mesh (BD Biosciences Pharmingen) to remove aggregates. The collected cells were washed, seeded on gelatin pre-coated 6-well plates and cultured in complete endothelial growth medium (EGM-MV; Lonza) supplemented with antibiotics. After 5–7 days, when cells reached confluency, a positive CD31 magnetic-bead selection was performed (CD31 MicroBead, 130-091-935, Miltenyi Biotech) according to the manufacturer's guidelines, and purified cells were further cultured in EGM medium.

Blood outgrowth ECs. Cells were established and cultured as previously described[26]. In brief, blood samples (obtained under protocol S57123 (Commission Medical Ethics of UZ/KU Leuven)) were diluted with PBS before Ficoll PaquePLUS (GE Healthcare) density-gradient centrifugation at 1,000*g* for 20 min at room temperature. The mononuclear cell layer was collected, washed with PBS and resuspended in EGM2 medium (PromoCell). Cells were plated in collagen-coated flasks and the medium was replaced every 2 days. From day 7 onwards, cells were checked for the formation of colonies, which were allowed to grow to approximately 1 cm^2. Blood outgrowth EC colonies were then trypsinized and subcultured.

HEK293T and HEPG2 cells. Cells (obtained from the ATCC) were grown in DMEM, supplemented with 10% FBS, 100 U ml^{-1} penicillin and 100 μg ml^{-1} streptomycin. When HEPG2 cells were compared directly to ECs in short-term stable isotope tracing experiments, they were incubated in exactly the same medium as the ECs to rule out possible bias arising from the difference in medium formulation. We did not perform authentication of the HEK293T and HEPG2 cells.

Mouse liver ECs. Cells were isolated from perfused healthy livers of control or *Glul*^{ECKO} mice. Before perfusion, the mice were anesthetized with Nembutal (60 mg kg^{-1}). Mice were perfused with 5 ml of a water-based perfusion buffer containing 1.7 M NaCl, 84 mM KCl, 120 mM HEPES and 1 mM NaOH followed by 5 ml of a PBS-based digestion buffer containing 0.1% collagenase II (Life Technologies), collagenase I (Life Technologies), 2 mM $CaCl_2$, 1% antibiotic-antimycotic (Life Technologies) and 10% FBS (Biochrom) at a perfusion rate of 1 ml min^{-1}. Perfusion was considered complete when the liver and mesenteric vessels were blanched and the desired amount of digestion buffer (≥5 ml) had passed through the circulatory system. Livers were dissected, placed into a 50 ml conical tube with 3 ml of digestion buffer and incubated at 37 °C for approximately 30 min, with regular shaking of the tubes every 5 min. After digestion, the tissue was homogeneously dissociated and the reaction was stopped with 10 ml of isolation buffer containing PBS + 0.1% BSA (Sigma-Aldrich). Subsequently, the cell suspension was filtered through a 100-μm cell strainer and cells were washed twice with isolation buffer. Finally, the ECs were isolated by magnetic-bead sorting with Dynabeads (CELLection Biotin Binder Kit, Life Technologies) coated with anti-mouse CD31 (eBioscience, Anti-Mouse CD31 Clone 390), according to the manufacturer's instructions. In brief, the cell suspension was incubated with the beads at room temperature for 30 min in HulaMixer Sample Mixer (Life Technologies). Next, CD31$^+$ ECs were collected by putting the tubes on a DynaMag-50 Magnet (Life Technologies) and removing the supernatant. The procedure was repeated twice to remove cell debris. Finally, cells were resuspended in EGM2 medium (PromoCell) and plated at the desired density on cell-culture plates pre-coated with 0.1% gelatin, and grown to confluency.

Mouse astrocytes. The cells were prepared as previously described with minor changes[27]. In brief, spinal cords were dissected from 13-day-old C57BL/6J mouse embryos. Meninges and dorsal root ganglia were removed and a single cell population was obtained by digestion with 0.05% trypsin in combination with gentle trituration. The cell suspension was layered on a 6.2% OptiPrep (Axis-Shield) cushion and centrifuged at 500*g* for 15 min. The pellet was resuspended and the cells were plated (12,000 cells per cm^2) in L15 medium supplemented with glucose (3.6 mg ml^{-1}), sodium bicarbonate (0.2%), penicillin (100 IU ml^{-1}), streptomycin (100 μg ml^{-1}) and FBS (10%). After reaching confluency, cell division was halted by treatment with cytosine arabinoside (10 μM, 3 days). After 4 weeks, more than 95% of cells stained positive for glial fibrillary acidic protein (not shown). We routinely tested primary cells and cell lines for mycoplasma contamination with the MycoAlert mycoplasma detection kit (Lonza, LT07-418).

Plasmid constructions and lentiviral particle production. cDNA for human *GLUL* was obtained from Origene. Silent mutations were introduced to make the *GLUL* cDNA resistant to the *GLUL*-specific shRNA (see below, TRCN0000045628). Constructs with point mutations were generated with Stratagene's QuickChange site-directed-mutagenesis kit following the manufacturer's guidelines. The cDNA for *RHOJ–eGFP* (*GFP–TCL*) was a gift from C. Der (Addgene plasmid 23231)[23] and was used as a template to generate the N-terminal-truncated ΔN20-RHOJ–eGFP, which lacked the first 20 amino acids, and Flag-tagged RHOJ. Standard cloning techniques were used to fuse GS to the photoswitchable fluorescent protein mEOS (pRSETa–mEos2 was a gift from L. Looger; Addgene plasmid 20341)[28]. The bimolecular fluorescence complementation (BiFC) approach vector enabling simultaneous expression of two separate cDNAs fused to eGFP subfragment 1 (N-terminal; containing amino acids 1 to 158) or subfragment 2 (C-terminal; containing amino acids from 159 onwards), respectively, was a gift from H. Mizuno (KU Leuven). GS was fused to the N-terminal subfragment of eGFP and RHOJ was fused to the C-terminal subfragment of eGFP to generate GS–eGFP$^{1/2}$, RHOJ–eGFP$^{2/2}$. Lentiviral expression constructs were obtained by cloning the respective cDNAs into pRRLsinPPT.CMV.MCS MM WPRE-vector. Validated *GLUL*-specific (TRC clones TRCN0000045628 (used in the majority of the experiments and indicated as *GLUL*^{KD1} in Extended Data Fig. 2a) and TRCN0000045631 (indicated as *GLUL*^{KD2} in Extended Data Fig. 2a and only used to confirm the migration and lamellipodial defect in Extended Data Fig. 3a, b) and *RHOJ*-specific (TRCN0000047606) shRNAs were either used in the pLKO.1 vector or subcloned into the pLVX-shRNA2 vector (PT4052-5; Clontech, Westburg BV). Scrambled shRNAs or the empty vectors were used as negative controls (both with the same outcome). All constructs were sequence-verified. Lentiviral particles were produced in HEK293T cells as previously described[2].

Recombinant protein production. Template vectors pRRLhGS, pRRLhGS^{R324C} and pRRLhGS^{R341C} containing the genes encoding wild-type *GLUL* or *GLUL* with point mutations were used as templates for PCR-based cloning. Recombinant constructs were expressed in the *Escherichia coli* strain BL21 codon + pICA2 that was transformed with pLH36-hGS in which expression is induced by isopropyl

β-D-1-thiogalactopyranoside under control of a pL-promotor developed by the Protein Core of VIB[29,30]. The pLH36 plasmid is provided with a His6-tag followed by a murine caspase-3 site. The murine caspase-3 site can be used for the removal of the His6-tag attached at the N terminus of the protein of interest during purification. The transformed bacteria were grown in 200 ml Luria Bertani medium supplemented with ampicillin (100 μg ml^{-1}) and kanamycin (50 μg ml^{-1}) overnight at 28 °C before 1/100 inoculation in a 20 l fermenter provided with Luria Bertani medium supplemented with ampicillin (100 μg ml^{-1}) and 1% glycerol. The initial stirring and airflow was 200 r.p.m. and 1.5 l min^{-1}, respectively. Further, this was automatically adapted to keep the pO_2 at 30%. The temperature was kept at 28 °C. The cells were grown to an optical density at 600 nm ($OD_{600\,nm}$) of 1.0, transferred at 20 °C, and expression was induced by addition of 1 mM isopropyl β-D-1-thiogalactopyranoside overnight. Cells were then collected and frozen at −20 °C. After thawing, the cells were resuspended at 3 ml g^{-1} in 50 mM HEPES pH 7.5, 500 mM NaCl, 20mM imidazole, 1 mM phenylmethylsulfonyl fluoride, 10% glycerol, 5 mM β-mercaptoethanol, 1 mg per 100 ml DNaseI (Roche) and 1 tablet per 100 ml Complete Protease Inhibitor (Roche). The cytoplasmic fraction was prepared using the Emulsiflex C3 (Avestin) followed by centrifugation. All steps were conducted at 4 °C. The clear supernatant was applied to a 20 ml Ni-Sepharose 6 FF column (GE Healthcare), equilibrated with 50 mM HEPES pH 7.5, 500 mM NaCl, 20 mM imidazole, 10% glycerol, 5 mM β-mercaptoethanol and 1 mM phenylmethylsulfonyl fluoride. The column was eluted with 50 mM HEPES pH 7.5, 500 mM NaCl, 400 mM imidazole, 10% glycerol, 5 mM β-mercaptoethanol and 1 mM phenylmethylsulfonyl fluoride after an intermediate elution step with 50 mM imidazole in the same buffer. Finally, the elution fraction was injected on a HiLoad 26/60 Superdex prep grade column with 20 mM HEPES pH 7.5, 300 mM NaCl, 10% glycerol and 0.5 mM tris(2-carboxyethyl)phosphine as running solution. The obtained elution fractions were analysed by SDS–PAGE. Recombinant protein concentration was determined using the Micro-BCA assay (Pierce).

In vitro knockdown and overexpression strategies. To minimize off-target effects and other silencing artefacts, key findings were confirmed with at least two independent and validated *GLUL*-specific shRNAs (see above) and appropriate controls or with a *GLUL*-specific siRNA duplex (5′-GGAAUAG CAUGUCACUAAAGCAGGC-3′) and scrambled control (TriFECTa, IDT). For lentiviral transduction of shRNAs or overexpressing constructs a multiplicity of infection (MOI) of 10 or 5 was used, respectively. In the case of simultaneous transduction of two different shRNAs, a MOI of 7.5 was used for each individual shRNA. In the case of simultaneous transduction of a shRNA in combination with an overexpression construct, the shRNA was transduced at a MOI of 10 and the overexpression construct at a MOI of 5, except for overexpression constructs for shRNA-resistant *GLUL* which were transduced at a MOI of 2.5. Transductions were performed on day 0 in the evening, cells were refed with fresh medium on day 1 in the morning and experiments were performed from day 3 or 4 onwards. siRNA transfection mixtures (in a total volume of 500 μl) were prepared in Opti-MEM containing GlutaMAX-I (Invitrogen) with Lipofectamine RNAi Max transfection reagent (Invitrogen) according to the manufacturer's instructions. The mixtures were added to the cells (150,000 cells in a 6-well-format plate) together with 2 ml EBM2 without antibiotics for overnight transfection, after which the medium was changed back to the regular M199 culture medium. siRNA transfection was performed at least 48 h before functional assays. BiFC plasmids were transfected into HEK293T cells with Fugene HD transfection reagent following the manufacturer's guidelines. Knockdown efficiency and overexpression levels were closely monitored for each experiment either on the mRNA (by PCR with reverse transcription; RT–PCR) or the protein level.

RNA isolation and gene expression analysis. Total RNA was extracted with PureLink RNA mini kit (Invitrogen) according to the manufacturer's instructions; quality and quantity were measured on a Nanodrop (Thermo Scientific). cDNA synthesis was performed with the iScript cDNA synthesis kit (BioRad). qPCR analyses were performed as previously described[1] on an Applied Biosystems 7500 Fast device with in-house-designed primers and probes or premade primer sets (Applied Biosystems or Integrated DNA Technologies) for which sequences and/or primer set ID numbers are available upon request. *ENOX2* or *HPRT1* were used as housekeeping genes.

Western blotting and (co-)immunoprecipitation. Proteins were extracted in Laemmli buffer (125 mM Tris-HCl (pH 6.8), 2% SDS, 10% glycerol) or in RIPA buffer (25 mM Tris-HCl (pH 7.6), 150 mM NaCl, 1% NP-40, 1% sodium deoxycholate, 0.1% SDS) containing protease and phosphatase inhibitor mixes (Roche Applied Science). After shearing of genomic DNA, proteins in the lysates were separated by SDS–PAGE, transferred to nitrocellulose or polyvinylidene difluoride membranes and detected with specific antibodies and HRP-conjugated secondary antibodies in combination with enhanced chemiluminescence or SuperSignal Femto western blotting substrate (Thermo Scientific). Signal was acquired with Image Quant LAS 4000 V 1.2 and densitometric quantification was carried out with ImageJ. For MLC and pMLC immunoblotting, each sample was loaded on two separate gels. One gel was used to detect MLC and the second was used to detect pMLC. Both gels had their own loading control, namely α-tubulin. pMLC/MLC was quantified as follows: (pMLC/α-tubulin)/(MLC/α-tubulin), abbreviated in the figure panel as (c)pMLC/(c)MLC with (c) meaning 'corrected for corresponding loading control'. Membrane versus cytosolic protein fractions were purified with the Plasma Membrane Protein Extraction Kit (101Bio) according to the manufacturer's guidelines and using proprietary buffers. For co-immunoprecipitation (co-IP) of endogenous or overexpressed proteins, ECs were lysed by rotating at 4 °C for at least 4 h in co-IP lysis buffer (20 mM Tris-HCl pH8, 137 mM NaCl, 10% glycerol, 1% nonidet NP-40 and 2 mM EDTA). Equal amounts of protein were incubated overnight with specific antibodies or matching isotype control IgGs at 4 °C. Subsequently, 20 μl of protein A/G–sepharose beads was added to the immune complexes for 4 h at 4 °C under gentle rotation. The beads were pelleted, washed three times with ice-cold co-IP lysis buffer and boiled for 5 min in reducing agent and loading buffer before SDS–PAGE. To determine the effect of deleting the first 20 N-terminal amino acids of RHOJ on its interaction with GS, co-IPs were carried out as above on ECs simultaneously overexpressing GS and RHOJ–eGFP or ΔN20-RHOJ–eGFP. In some of the experiments, the expression of the ΔN20-RHOJ–eGFP was lower than the expression of RHOJ–eGFP. To correct for this possible bias, densitometric quantification of all bands was performed in ImageJ and signals in the immunoprecipitation lanes were normalized to the input signals. The amount of GS immunoprecipitated was the same under the RHOJ–eGFP and ΔN20-RHOJ–eGFP conditions (data not shown).

Biochemical and metabolic assays. *Bicinchoninic acid assay.* The BCA protein assay kit (Pierce) was used to determine protein content with Gen5 1.11.5 (BioTek Instruments).

Lactate dehydrogenase release. This was determined as a measure of cell survival using the Cytotoxicity Detection Kit (Roche Applied Science) with Gen5 1.11.5 (BioTek Instruments).

Intracellular reactive oxygen species. Levels were determined by CM-H_2DCFDA dye (5-(and-6)-chloromethyl-2′,7′-dichlorodihydrofluorescein diacetate, acetyl ester; Invitrogen) labelling following the manufacturer's guidelines.

Glutamine synthetase activity. The enzyme activity in living cells was determined by pulse-labelling the cells for 30 min with 2 mM $^{15}NH_4Cl$ and subsequent determination of ^{15}N incorporation in intracellular glutamine by gas chromatography–mass spectrometry (GC–MS). Similarly, GS activity was measured by pulse-labelling for 30 min with 0.5 mM [U-^{13}C]glutamic acid and subsequent tracing of ^{13}C into glutamine by GC–MS. The 0.025 mM glutamine condition was added to this assay for the sole purpose of having a positive control—lowering external glutamine levels should increase GS activity—and does not in any way reflect maximal GS activity. Background signals were determined by pre-incubating the cells with the GS inhibitor MSO. As an independent method (not relying on labelling one of the immediate substrates (NH_4^+ or glutamate)) of determining GS activity, we performed steady-state labelling of ECs with [U-^{13}C]glucose (5.5 mM) and determined the ^{13}C contribution to α-ketoglutarate, glutamate and glutamine (for labelling scheme, see Extended Data Fig. 5f). Before derivatization for GC–MS analysis, cells were washed with ice-cold 0.9% NaCl and extracted in ice-cold 80/20 methanol/water.

Glutamine uptake assay. Dynamic [U-^{13}C]glutamine uptake assays were performed as follows: 2.5×10^5 cells per well were seeded in 6-well plates and pulse-labelled for 0, 0.5, 10, 20 and 30 min with the regular M199 culture medium containing 0.6 mM [U-^{13}C]glutamine instead of the regular 0.6 mM unlabelled glutamine. The 0-min time point represents an absolute negative control for which extracts were made from ECs that had never been treated with tracer-containing medium. For the 0.5-min time point, the labelled medium was put on the cells and immediately aspirated (altogether taking 0.5 min). At all time points, cells were thoroughly washed twice with ice-cold 0.9% NaCl to ensure complete removal of tracer-containing medium. Cellular extracts were then made in ice-cold 80/20 methanol/water, before derivatization for GC–MS measurements. Alternatively, cells were incubated with 0.5 μCi ml^{-1} [U-^{14}C]-L-glutamine for 10 min, after which they were washed at least three times with ice-cold PBS. The last PBS wash was collected and checked for residual radioactivity. Cells were then lysed with 200 μl 0.2 M NaOH and lysates were neutralized with 20 μl 1 M HCl and used for scintillation counting.

[^{3}H]thymidine incorporation. Proliferation was determined by labelling the cells with 1 μCi ml^{-1} [^{3}H]thymidine for 2 h, followed by fixation in 100% ethanol for 15 min, precipitation with 10% trichloroacetic acid and finally lysis in 0.1 M NaOH. Scintillation counting was used to assess the amount of [^{3}H]thymidine incorporated into the DNA.

Energy charge assessment. Cells (1.5×10^6) were collected in 100 μl ice-cold 0.4 M perchloric acid containing 0.5 mM EDTA. The pH was adjusted with 100 μl of 2 M K_2CO_3. One hundred microlitres of the mixture was subsequently injected onto an Agilent 1260 HPLC with a C18-Symmetry column (150 × 4.6 mm, 5 mm; Waters), maintained at 22.5 °C. The flow rate was kept constant at 1 ml min^{-1}. A linear gradient using solvent A (50 mM NaH_2PO_4, 4 mM tetrabutylammonium, adjusted to

pH 5.0 with H_2SO_4) and solvent B (50 mM NaH_2PO_4, 4 mM tetrabutylammonium, 30% CH_3CN, adjusted to pH 5.0 with H_2SO_4) was accomplished as follows: 95% A for 2 min, from 2 to 25 min linear increase to 100% B, from 25 to 27 min isocratic at 100% B, from 27 to 29 min linear gradient to 95% A and finally from 29 to 35 min at 95% A. ATP, ADP and AMP were detected at 259 nm.

Seahorse extracellular flux measurements. ECs were seeded at 1.5×10^5 cells per well on Seahorse XF24 tissue culture plates (Seahorse Bioscience Europe). Oxygen consumption (OCR) measurements were performed at 6-min intervals (2 min mixing, 2 min recovery, 2 min measuring) in a Seahorse XF24 device (XF Reader 1.8.1.1 software). Consecutive treatments with oligomycin (1.2 μM final), FCCP (5 μM final) and antimycin A (1 μM final) were performed to enable quantification of ATP-coupled OCR (OCR_{ATP}) and maximal respiration, next to basal OCR (OCR_{bas}).

Glycolytic flux. ECs were cultured for 6 h in medium containing 0.4 μCi ml^{-1} [5-^{3}H]D-glucose (Perkin Elmer) after which the supernatant was transferred into glass vials sealed with rubber stoppers. ^3H_2O was captured in hanging wells containing a Whatman paper soaked with H_2O over a period of 48 h at 37 °C to reach saturation[1]. The paper was then used for liquid scintillation counting (QuantaSmart V4 Perkin Elmer).

[*^{14}C*]*glucose oxidation.* ECs were incubated for 6 h in medium containing 0.55 μCi ml^{-1} [6-^{14}C]D-glucose. After that, 250 μl of 2 M perchloric acid was added to each well to stop cellular metabolism and to release $^{14}CO_2$, which was captured overnight at room temperature in 1X hyamine hydroxide-saturated Whatman paper. The radioactivity in the paper was determined by liquid scintillation counting (QuantaSmart V4 Perkin Elmer)[1].

[*^{14}C*]*glutamine oxidation.* ECs were incubated for 6 h with medium containing 0.5 μCi ml^{-1} [U-^{14}C]glutamine. Two hundred and fifty millilitres of 2 M perchloric acid was added to the cells to stop cellular metabolism and release $^{14}CO_2$. Trapping of $^{14}CO_2$ occurred as described for ^{14}C-glucose oxidation[1].

Protein (auto)palmitoylation detection. *In vitro palmitoylation (click-reaction-based).* Purified bacterial GS protein was incubated with the indicated concentration of palmitoyl alkyne-CoA (Cayman Chemical) for 6 h at room temperature. The GS protein was then denatured by the addition of SDS. A click reaction with biotin azide was performed to label the palmitoylated proteins[25]. Palmitoylated proteins were detected by SDS–PAGE followed by blotting with streptavidin–horseradish peroxidase.

Fluorescence-based CoA release detection. During autopalmitoylation of proteins, palmitate is transferred from palmitoyl-CoA to the protein, thereby releasing reduced CoA. α-Ketoglutarate dehydrogenase can use CoA to convert α-ketoglutarate to succinyl-CoA, a reaction that features the reduction of NAD^+ to fluorescent NADH[31]. In brief, recombinant human GS was incubated with palmitoyl-CoA in MES buffer at physiological pH for at least 1 h at 30 °C. The volume was then adjusted to 200 μl in 50 mM sodium phosphate buffer (pH 6.8) containing 2 mM α-ketoglutaric acid, 0.25 mM NAD^+, 0.2 mM thiamine pyrophosphate, 1 mM EDTA, 1 mM dithiothreitol and 32 mU α-ketoglutarate dehydrogenase. NADH levels were measured at 20 min after initiation of the reaction on a VICTOR plate reader (340 nm excitation, 465 nm emission). The experiment was performed in two directions: either with varying doses of palmitoyl-CoA for a fixed amount of recombinant GS (2 μg) or with varying amounts of recombinant GS for a fixed concentration of palmitoyl-CoA (40 μM).

Affinity chromatography. A previously published protocol was used to determine cell-free binding of recombinant human GS to palmitoyl-CoA agarose[32]. A total of 50 μl of immobilized palmitoyl-CoA-agarose was equilibrated with 20 mM Tris-HCl (pH 8.4)/120 mM NaCl buffer. The beads were incubated with 40 μg of recombinant human GS in a final volume of 200 μl for 2 h at room temperature on a rotatory system. Beads were pelleted and 20 μl of the supernatant was collected as the flow-through fraction. Beads were then washed eight times with 500 μl of 20 mM Tris-HCl (pH 8.4)/120 mM NaCl buffer. Twenty microlitres of the last wash fraction was collected as fraction wash-8. Beads were then eluted with SDS loading buffer and heated for 15 min at 60 °C. Two micrograms of recombinant protein was used as input fraction. Input fraction, flow-through, wash-8 and SDS-eluate were analysed by immunoblotting for the presence of GS.

In-cell labelling. In-cell labelling experiments were performed essentially as previously described[25]. HEK-293T cells were transfected with the indicated expression plasmids. Twenty-four hours after transfection, the medium was replaced with DMEM + 10% dialysed FBS containing the indicated probes (50 μM 16C-YA or 50 μM 16C-BYA). After 18 h, cell lysates were collected by incubation of the cells on ice for 15 min in lysis buffer (50 mM TEA-HCl (pH = 7.4), 150 mM NaCl, 1% Triton X-100, 0.5% sodium deoxycholate, 0.1% SDS and 5 mM phenylmethanesulfonyl fluoride) followed by centrifugation for 10 min at 15,000*g*. Equal amounts of protein were then used for a click reaction with biotin azide. For labelling with 17-ODYA, Flag–RHOJ overexpressing ECs were incubated overnight with 17-ODYA (50 μM) in M199 supplemented with 3.6% fatty-acid-free BSA, 10% dialysed FBS and 5 mM sodium pyruvate. Cells were washed with ice-cold PBS and lysed in NaP lysis buffer (0.2 M $Na_2HPO_4{\cdot}2H_2O$, 0.2 M $NaH_2PO_4{\cdot}2H_20$, 1 M NaCl, 10% NP40). Two micrograms of anti-Flag antibody was conjugated to 20 μl of Dynabeads protein G (Thermo Fisher) for 1 h at room temperature. After washing the beads twice with NaP lysis buffer, at least 500 μg of protein was added to the beads for 3 h at 4 °C. Then beads were washed three times with NaP lysis buffer and resuspended in 20 μl of resuspension buffer (4% SDS, 50 mM TEA, 150 mM NaCl). The click reaction was initiated by adding 0.5 μl of 5 mM TAMRA-azide (Lumiprobe), 0.5 μl 50 mM tris(2-carboxyethyl)phosphine hydrochloride (TCEP-HCl), 0.5 μl 10 mM tris[(1-benzyl-1*H*-1,2,3-triazol-4-yl)methyl]amine (TBTA) and 2.4 μl of 5 mM freshly prepared ascorbic acid. Samples were then incubated for 1 h at 37 °C in the dark. Sample buffer (9.4 μl) and reducing agent (3.7 μl) were added to stop the reaction. After 10 min at room temperature in the dark, samples were frozen at −80 °C or run on a 10% Bis-Tris gel in MES buffer. In-gel fluorescence was imaged with Typhoon FLA 9500 V1.0.

Streptavidin pull-down. After click reaction with biotin azide, free biotin azide was removed from the samples by centrifugal filtration column (Millipore). The samples were then incubated with streptavidin-conjugated beads for 1 h at room temperature. After washing with PBS-T, proteins were eluted from the beads by incubation in elution buffer (95% formamide, 10 mM EDTA (pH = 8.0)) at 95 °C for 5 min.

Acyl-resin-assisted capture. This method, in which free cysteine thiols are chemically blocked and palmitoylated cysteines are exposed and captured by a resin, was performed with the CAPTUREome S-Palmitoylated Protein Kit (Badrilla) with minor adaptations to the manufacturer's guidelines. Five hundred micrograms of protein were incubated for 4 h in 500 μl of thiol blocking reagent (to block free thiols). Proteins were precipitated with ice-cold acetone and afterwards solubilized with 300 μl of binding buffer and spun down. After protein quantification, 30 μg was kept as total input fraction (IF), and equal amounts of protein were incubated for 2.5 h with or without (to obtain the negative control preserved bound fraction pBF) a thioester linkage specific cleavage reagent to cleave the thioester bond. Newly liberated thiols were captured with CAPTUREome resin. The resin was spun down and the supernatant was collected as the cleaved unbound fraction (cUF) to check whether the proteins of interest were indeed completely depleted from the thioester cleavage reagent (meaning efficient capture of the free thiols by the resin). After thorough washing of the resin, captured proteins (cleaved bound fraction, cBF), were eluted with reductant and analysed together with the input fraction, cleaved unbound fraction and preserved bound fraction by SDS–PAGE followed by immunoblotting.

GC–MS analysis. Metabolites from cells were extracted in 800 μl 80% methanol (at −80 °C). Next the extracts were centrifuged at 4 °C for 15 min at 20,000*g* and the supernatants were dried in a vacuum centrifuge. Twenty-five microlitres of a 2% methoxyamine hydrochloride solution (20 mg dissolved in 1 ml pyridine) was added to the dried fractions, which were then incubated at 37 °C for 90 min. Then 75 μl of *N-tert*-butyldimethylsilyl-*N*-methyltrifluoroacetamide with 1% *N-tert*-butyldimethylchlorosilane (Sigma-Aldrich) was added and the reaction was carried out for 30 min at 60 °C. Reaction mixtures were centrifuged for 15 min at 20,000*g* at 4 °C to remove insolubilities and the supernatant was transferred to a glass vial with conical insert (Agilent). GC–MS analyses were performed on an Agilent 7890A GC equipped with a HP-5 ms 5% Phenyl Methyl Silox (30 m, 0.25 mm internal diameter, 0.25 μm; Agilent Technologies) capillary column, interfaced with a triple quadrupole tandem mass spectrometer (Agilent 7000B, Agilent Technologies) operating under ionization by electron impact at 70 eV. The injection port, interface and ion source temperatures were kept at 230 °C. The temperature of the quadrupoles was kept at 150 °C. The injection volume was 1 μl, and samples were injected at a 1:10 split ratio. The helium flow was kept constant at 1 ml min^{-1}. The temperature of the column started at 100 °C for 5 min and increased to 260 °C at 2 °C min^{-1}. Next, a 40 °C min^{-1} gradient was applied until the temperature reached 300 °C. After the gradient, the column was heated for another 3 min at 325 °C. The GC–MS analyses were performed in single-ion-monitoring scanning mode for the isotopic pattern of metabolites.

LC–MS analysis. Polar metabolites were extracted using 250 μl of a 50/30/20 methanol/acetonitrile/10 mM ammonium acetate pH 9.3 (containing 2 μM of deuterated (d27) myristic acid as internal standard) extraction buffer. Following extraction, precipitated proteins and insoluble matter were removed by centrifugation at 20,000*g* for 20 min at 4 °C. The supernatant was transferred to the appropriate mass spectrometry vials. Measurements were performed using a Dionex UltiMate 3000 LC System (Thermo Scientific) in-line connected to a Q-Exactive Orbitrap mass spectrometer (Thermo Scientific). Fifteen microlitres of sample was injected and loaded onto a Hilicon iHILIC-Fusion(P) column (Achrom). A linear gradient was applied starting with 90% solvent A (LC–MS grade acetonitrile) and 10% solvent B (10 mM ammonium acetate pH 9.3). From 2 to 20 min the gradient changed to 80% B and was kept at 80% until 23 min. Next a decrease to 40% B was applied to 25 min, further decreasing to 10% B at 27 min. Finally, 10% B was maintained until 35 min. The solvent was used at a flow rate of 200 μl min^{-1}, the

column temperature was kept constant at 25 °C. The mass spectrometer operated in negative-ion mode, and the settings of the HESI probe were as follows: sheath gas flow rate at 35, auxiliary gas flow rate at 10 (at a temperature of 260 °C). Spray voltage was set at 4.8 kV, temperature of the capillary at 300 °C and S-lens RF level at 50. A full scan (resolution of 140,000 and scan range of m/z 70–1,050) was applied. For the data analysis, we used an in-house library and metabolites of interest were quantified (area under the curve) using the XCalibur 4.0 (Thermo Scientific) software platform.

In vitro assays. *Endothelial spheroid capillary sprouting.* The assay was performed following established protocols[1,2]. To form the spheroids, ECs were cultured overnight in hanging drops in EGM2 medium with methylcellulose (Sigma-Aldrich; 20% volume of a 1.2% solution of methylcellulose viscosity 4,000 cP). Spheroid sprouting involves both EC proliferation and migration. To have a 'clean' view on the migration aspect in sprouting, we also included conditions in which we blocked EC proliferation before sprout formation. More precisely, mitotic inactivation was achieved by adding mitomycin C (1 μg ml^{-1}) to the medium. To induce sprouting, spheroids were embedded in a collagen gel and incubated for 20 h. If required, chemical compounds (Fasudil at 10 μM, H1152 at 1 μM and Y26732 at 10 μM) were added during the collagen-gel-incubation step. Spheroids were then fixed with 4% paraformaldehyde and imaged under phase-contrast illumination with a Motic AE 31 microscope (Motic Electric Group) or a Leica DMI6000B microscope (Leica Microsystems). Phase-contrast images were used to quantify the number of sprouts per spheroid and the total sprout length (cumulative length of all sprouts on a spheroid). Spheroid body circumference was measured to correct for differences in sizes of the spheroid. Per experiment (that is, per individual HUVEC isolation) at least 10 spheroids per condition were analysed.

Scratch wound assays. Seventy-five thousand HUVECs were seeded in a 24-well format and were allowed to reach confluency over the next 24 h. At time T_0 the confluent monolayer was scratched with a 200-μl pipette tip and photographed. The cells were further incubated for the indicated times and photographed again at time point T_x. The gap area at T_0 minus the gap area at T_x was measured with ImageJ and expressed as % migration distance. Per well, three non-overlapping regions along the scratch were analysed. Much like the spheroid sprouting, scratch wound healing is a combined readout for EC migration and proliferation. Therefore, we also included conditions in which the ECs were pre-treated with mitomycin C (1 μg ml^{-1}) to rule out the effect of proliferation.

Boyden chamber assays. Fifty thousand HUVECs were seeded on 0.1% gelatin-coated transwells and allowed to adhere. Then, the transwells were washed and refed with medium containing only 0.1% FBS and placed in bottom wells containing medium with 5% FBS as a pro-migratory stimulus. Sixteen hours later, transwells were processed and analysed for numbers of migrated cells. Pre-treatment with mitomycin C (see above) was applied.

Velocity of random movement. This was assessed on HUVECs that were sparsely seeded on glass-bottom 24-well plates. Time-lapse videos were generated by confocal image acquisition at 4-min intervals. Velocity of movement was determined by tracking the nucleus position as a function of time (μm h^{-1}) (Tracking Tool, Gradientech AB). Per condition, on average 2 or 3 individual cells were traced in each biological repeat.

Lamellipodial area. This was measured on sparsely seeded phalloidin-stained ECs with Leica MM AF morphometric analysis software (Leica Microsystems) with in-house-developed macros and is expressed in percentage of total cell area. Treatment with MSO (1 mM), Y27632 (10 μM), fasudil (10 μM), H1152 (1 μM), ML7 (15 μM) and peptide 18 (15 μM) was carried out 24 h before analysis of the cells. Per experimental condition, a minimum of ten individual cells was analysed.

Staining and quantification of VE-cadherin junctions. VE-cadherin staining and quantification of junctional length and gap index was performed as previously described[33]. First, the total junctional length (100%) was determined by summing up all segments, then the sum of all continuous segments was calculated as the percentage of total junctional length. The percentage difference between total and continuous represents the discontinuous length. Gap size index (intercellular gap area/cell number) was determined with the formula ([intercellular gap area/total cell area] × 1,000)/cell number. Junctional lengths, intercellular gap area and total cell area were defined manually with ImageJ. For each condition, a minimum of 10 fields was quantified (10–15 cells per field on average) per experiment, and data shown represent the mean of at least three independent experiments.

Transendothelial electrical resistance. Fifty thousand ECs were seeded on 6.5-mm 0.1% gelatin-coated polyester transwells, 0.4-μm pore size (Costar ref. 3470, Sigma-Aldrich). The electrical resistance was measured with an Endhome-6 electrode (World Precision Instruments) connected to an EVOM2 voltohmmeter (World Precision Instruments). Gelatin-coated wells without cells were used to measure the intrinsic electrical resistance of the inserts for background subtraction. Measurements were performed every day for four consecutive days, with at least two measurements per condition.

Actin dynamics and Rho (kinase) activity assays. *Latrunculin wash-out.* ECs were treated with latrunculin B (100 ng ml^{-1}) for 30 min and were then washed three times with culture medium. The cells were fixed at the indicated time points and stained with phalloidin to visualize actin stress fibres.

The F- actin/G-actin ratio. The ratio was determined for $GLUL^{KD}$ versus control ECs, in 4% paraformaldehyde-fixed cells that were permeabilized for 10 min in PBS with 0.2% Triton X-100 and stained with phalloidin-Alexa 488 and DNase I-Alexa 594 (1:200)[34]. Fluorescence intensities were quantified with ImageJ and were based on grey values. On average, ten individual cells were analysed per experimental condition.

RHOJ activity measurements. Cells were lysed in buffer containing 50 mM Tris, pH 7.6, 150 mM NaCl, 1% Triton X-100, 0.5 mM $MgCl_2$, protease inhibitors and 0.1 μg μl^{-1} biotinylated CRIB-peptide. After spinning down for 4 min at 18,000*g* at 4 °C, 50 μl streptavidin-coated beads was added to the lysates. Subsequently, samples were rotated for 30 min at 4 °C, beads were washed 4 times in the above buffer after which they were boiled for 5 min in reducing agent and loading buffer[35]. As negative controls in this assay, we used lysates from $RHOJ^{KD}$ ECs, a streptavidin-beads only-condition and lysates in which the biotinylated CRIB-peptide was replaced by an irrelevant biotinylated protein (Fig. 4c).

RHOA/B/C activity. This was determined with glutathione *S*-transferase–Rhotekin pull down assays following previously established protocols[36].

ROCK activity. This was assayed by determining phosphorylation of the ROCK target myosin light-chain 2 on western blot or by immunostaining. Fluorescence intensities from immunostaining were quantified with ImageJ and were based on grey values.

Confocal and high-resolution imaging. *Confocal imaging.* Imaging was performed on a Zeiss LSM 510 Meta NLO or Zeiss LSM 780 confocal microscope (oil objectives: ×40 with numerical aperture (NA) 1.3, ×63 with NA 1.4, ×100 with NA 1.3) with ZEN 2011 software (Carl Zeiss). Within individual experiments, all images across different experimental conditions were acquired with the same settings.

DORA–RHOA–FRET imaging. RHOA activity was measured in living HUVECs by monitoring fluorescence intensities of the FRET-acceptor yellow fluorescent protein (YFP) and the donor cyan fluorescent protein (CFP) as previously described[37]. In brief, a Zeiss Observer Z1 microscope, with a Chroma 510 DCSP dichroic splitter, two Hamamatsu ORCA-R2 digital CCD cameras and an attached dual camera adaptor (Zeiss) controlling a 510 DCSP dichroic mirror, was used for simultaneous monitoring of CFP and YFP emissions using filter sets ET 480/40 and ET 540/40m (Chroma Technology), respectively. To excite the CFP donor, ET 436/20x and 455 DCLP dichroic mirror was used (Chroma). For FRET/CFP ratiometric processing, CFP and YFP images were processed using the MBF ImageJ collection. The images were background-subtracted, aligned and a threshold was applied. Finally, the FRET/CFP ratio was calculated and a custom lookup table was applied to generate a colour-coded image, in which white and red colours illustrate high and blue colours illustrate low RHOA activities.

BiFC imaging and quantification. BiFC was evaluated using a laser scanning microscope (Fluoview FV1000, Olympus) equipped with a UPLSAPO 60× oil objective (NA 1.35). Before imaging, cells were fixed with 4% (v/v) paraformaldehyde and stained with 4′,6-diamidino-2-phenylindole (DAPI; 1:1,000 dilution, Invitrogen). A 488-nm laser was used for exciting eGFP, whereas DAPI was excited using a 405-nm laser. A DM405/488/559/635 polychroic mirror was used to guide the excitation lasers to the sample. Fluorescence images of fixed cells were acquired using a sampling speed of 4 μs per pixel. Emission light was collected at 430–470 nm and 500–550 nm, for DAPI and eGFP, respectively. The images were acquired with a pixel size of 207 nm (1,024 × 1,024 pixels). BiFC was first established in HEK cells expressing GS–eGFP$^{1/2}$ and RHOJ–eGFP$^{2/2}$ from one expression vector, with a construct overexpressing an unfused N-terminal eGFP half-site together with RHOJ coupled to the C-terminal eGFP half-site as a negative control (data not shown). To determine the effect of deleting the first 20 amino acids in RHOJ on BiFC in ECs, separate expression constructs for GS–eGFP$^{1/2}$, RHOJ–eGFP$^{2/2}$ and ΔN-RHOJ–eGFP$^{2/2}$ were used (Extended Data Fig. 7e). Quantification of expression efficiency was carried out using an in-house-generated routine in MATLAB.

TIRF microscopy. An in-house-generated setup based on an inverted microscope (IX83, Olympus) was used to detect single molecules under total internal reflection (TIRF) mode. The setup was equipped with an Electron Multiplying-CCD cameras (ImagEM C9100-13; Hamamatsu Photonics) and an APON 60XOTIRF objective lens (NA 1.49, Olympus). The GS–mEos3.2 molecules were excited with a 561-nm line from a DPSS laser (200 mW; Coherent) and converted with a 405-nm line from a diode laser (Cube, 100 mW; Coherent). Before expansion, the laser lines were combined using a 405bcm dichroic mirror. The laser lines were guided onto the sample by a dichroic mirror, z488/561/633rpc. The red fluorescence of the mEos3.2 was detected through a long-pass filter 572 (HQ572LP), in combination with a band-pass filter HQ590M40-2P. All the filters were purchased from Chroma. Time-lapse fluorescence images were recorded with continuous illumination at a 62.5-Hz acquisition rate (16 ms per frame).

Single-particle tracking. For calculation of single-molecule coordinates, the program 'Localizer' running from MATLAB was used[38]. After localization, the positions of a molecule detected in consecutive frames are connected to reconstruct a trajectory using in-house-developed software in MATLAB. Coordinates presented in consecutive frames are linked to form a single trajectory when they uniquely appear at a distance smaller than 856 nm (corresponding to 8 pixels). Trajectories with at least three steps were analysed using variational Bayes single-particle tracking analysis (vbSPT), a software package for analysis of single-particle diffusion trajectories, in which the diffusion constants switch randomly according to a Markov process[39].

For experiments involving microscopic analysis of individual (immuno-)stained cells, initial confocal image acquisition was done by researchers blinded to the experimental condition to avoid biased imaging of cells presenting a certain phenotype.

Mice. *$Glul^{ECKO}$ mice.* To obtain inducible EC-specific GS-knockout mice, *$Glul^{lox/lox}$* mice[40] were intercrossed with *VE-Cadherin-cre^{ERT2}* mice[41] or with *Pdgfb-cre^{ERT2}* mice[42] and named *$Glul^{iVECKO}$* and *$Glul^{iPECKO}$*, respectively. Correct Cre-mediated excision of the *loxP*-flanked *Glul* segment in tamoxifen-treated *$Glul^{ECKO}$* mice was confirmed by PCR analysis of genomic DNA (Extended Data Fig. 1d, e).

Generation of $Glul^{+/GFP}$ chimaeras. Blastocysts were collected from superovulated C57BL/6 females at post-coital day 3.5 and were cultured for 5–8 days in ES cell culture medium consisting of Knockout DMEM medium (Invitrogen), with 2 mM L-glutamine, FBS (Hyclone, Thermo Scientific), MEM non-essential amino acids 100X (Invitrogen), 0.01 mM β-mercaptoethanol (Sigma-Aldrich), 1 mM sodium pyruvate (Invitrogen), 100 U ml^{-1} penicillin, 100 μg ml^{-1} streptomycin and 2,000 U ml^{-1} leukaemia inhibitory factor (Merck, Millipore). Afterwards, the inner cell mass was selectively removed from the trophectoderm, trypsinized and replated on a mitomycin C-arrested MEF feeder monolayer. ES cells were fed every day and passaged every 2–4 days onto new feeder cells. *$Glul^{+/GFP}$* ES cells (E14IB10 ES cell line)[6] were injected into C57BL/6 blastocysts and high chimaeric pups were killed at P5 for detection of GFP in the retinal microvasculature.

In vivo models. *Analysis of dorsal dermal blood vessel network.* From E11.5 to E13.5 (where detection of the vaginal plug was designated E0.5), *$Glul^{iVECKO}$* pregnant dams were treated with tamoxifen (50 mg kg^{-1}) by oral gavage. At E16.5 they were euthanized by cervical dislocation after which embryos were dissected from the uterus. Yolk sacs were collected, washed with PBS and used for genotyping of the embryos. The embryos were fixed for 10 min in 1% paraformaldehyde before dissection of the dorsal skin. The epidermal and dermal layers were separated under a dissection microscope. Dissected back skins were permeabilized overnight (0.5% Triton X-100, 0.01% sodium deoxycholate, 1% bovine serum albumin, 0.02% sodium azide) before whole-mount immunostaining with CD31. To systematically analyse the same region for each embryo, 1 rectangular confocal image (1,700 × 1,100 μm) was taken at the anterior side of the skin specimen with the upper longer side of the rectangle placed on the midline. Within each rectangular picture the number of branch points was determined with the cell counter tool in ImageJ in 6 regions of interest (ROIs) (250 × 250 μm), three in the top half and three in the bottom half of the rectangle, not overlapping with the larger arteries and veins.

Neonatal retinal angiogenesis. EC-specific *Glul* deletion was obtained by intraperitoneal administration of tamoxifen (Sigma; 10 mg kg^{-1}; dissolved in 1:10 ethanol: oil solution) once daily from P1 to P3 in *$Glul^{iVECKO}$* or once at P2 for *$Glul^{iPECKO}$*. For in vivo proliferation quantification, EdU (Invitrogen) was injected intraperitoneally 2 h before euthanasia. Unless stated otherwise, retinas were isolated at P5 as previously described[43] and fixed in 2% paraformaldehyde for 2 h. Isolectin B4 (IB4), EdU, NG2 and ColIV stainings were performed as previously described[1,2]. Radial outgrowth of the vascular plexus, vascular area, branch points, number of filopodia and number of distal sprouts were analysed on IB4-stained retinas (see below) with ImageJ. Numbers of branch points and EdU$^+$ ECs were quantified in 200 × 200-μm ROIs; per retina 12 ROIs were placed at the front of the vascular plexus and 8 ROIs were placed more towards the centre of the plexus. Filopodia and distal sprouts were quantified on ten high-magnification (63×) images per retina, each representing approximately 200 μm of utmost vascular front. For analysis of the retinal vasculature at P21 (3 weeks old) and P42 (6 weeks old), mice underwent the same tamoxifen treatment regimen as for analyses at P5. In addition, different tissues were collected from P42 mice for endoglin and CD34 staining to study blood vessels in different vascular beds.

Oxygen-induced retinopathy. Oxygen-induced ROP was induced by exposing C57BL/6 pups to 70% oxygen from P7–P12. Pups were then returned to normoxia and injected daily with 20 mg kg^{-1} MSO. At P17, pups were euthanized and eyes were enucleated, fixed in 4% paraformaldehyde and retinal flatmounts were stained for isolectin B4[2,3]. MSO-treated mice retained normal behaviour notwithstanding observable weight loss. Mosaic tile images were captured using the inverted Leica DMI6000B epifluorescence microscope (Leica) and analysis of the vascular tuft area (the complete retina was analysed, no ROIs were used) and the vaso-obliterated area was performed with ImageJ software and are expressed as percentage of the total retinal area.

Corneal (micro-)pocket assay. This assay, to induce neovascularization of the avascular cornea, was performed as previously described[44]. In brief, in the eyes of eight-week-old C57BL/6 mice, a lamellar micropocket was dissected towards the temporal limbus to enable the placement of a basic fibroblast growth factor (bFGF)-containing pellet on the corneal surface. Five days after implanting the pellets, the mice were euthanized, the eyes were enucleated and the corneas were excised and fixed in 70% ethanol before CD31 antibody staining. After staining, the corneas were flat-mounted and imaged on a Zeiss LSM 780 confocal microscope. CD31$^+$ area was measured in ImageJ after thresholding the signal and is expressed as percentage of total cornea area. Production of the pellets was carried out as previously described[44]. The pellets contained 20 ng bFGF and the concentration of MSO in the initial solution from which the pellets were made was 10 mM.

Imiquimod-induced skin inflammation. Ten-week-old female BALB/C mice received a daily topical dose of 5% imiquimod cream (62.5 mg) on their shaved backs for four days to induce skin inflammation[3]. One hour after each administration of the cream, the same skin area was treated either with Vaseline jelly or Vaseline jelly containing MSO (low dose, 20 mg kg^{-1}; or high dose, 40 mg kg^{-1}). The MSO treatment did not affect the body weight of the mice. Skins were collected and fixed in 4% paraformaldehyde. Paraffin sections of skins were stained for CD105 (R&D Systems) and haematoxylin and eosin. Images were captured with a Leica DMI6000B microscope (Leica Microsystems). Per mouse, ten images representing different locations along the total length of the skin specimen were analysed for CD105$^+$ area.

Miles vascular permeability assay. Eight-week-old female BALB/C mice were treated for three consecutive days with 20 mg kg^{-1} day^{-1} MSO or with vehicle before injection with 300 μl 0.5% Evans blue dye. The inflammatory irritant mustard oil (0.25 ml allyl isothiocyanate in 4.75 ml mineral oil) was applied on one of the ears with a cotton swab to induce vascular permeability. Mineral oil as a control was applied on the other ear. After 15 min, mustard oil or mineral oil was again applied on the ear for 30 min, after which the circulation was flushed with saline for 3 min under complete anaesthesia and mice were perfused with 1% paraformaldehyde in 50 mM citrate buffer (pH 3.5) for 2 min. Ears were cut and minced in formamide and incubated at 55 °C overnight to extract the Evans blue from the tissue. Quantification of the dye was performed by a spectrophotometrical optical density measurement at 620 nm.

Haematological profiling in six-week-old mice. Profiling was performed with a Cell Dyn 3700 device (Abbott Diagnostics) according to the manufacturer's guidelines. Plasma measurements for different liver and inflammation parameters were performed in the clinical laboratory of UZ Leuven. Prior randomization was not applicable for any of the above mouse models given that all animal treatments were performed in baseline conditions. No statistical methods were used to predetermine the sample size. For all mouse experiments, data analysis was carried out by researchers blinded to the group allocation. All animal procedures were approved by the Institutional Animal Care and Research Advisory Committee of the University of Leuven.

In silico screening for palmitoylation sites. The human RHOJ protein sequence was screened for putative palmitoylation sites on the SwissPalm website[22] entering 'RHOJ' as the protein name.

Modelling and simulations. The GS models were built starting from X-ray crystallographic structures retrieved from the Protein Data Bank (entry 2OJW for human GS and 1FPY for bacterial GS). All simulations were run with Gromacs 5.1.4[45] and the Amber FF14SB[46] force field, while palmitoyl-CoA was parametrized with GAFF and the point charges were calculated with Gaussian 09[47] at the Hartree–Fock level with a 6-31G* basis set. The different models were then embedded in a TIP3P water box, and counter ions were added to ensure overall charge neutrality. An initial 2,000 steps of steepest descent and 500 steps of conjugated gradient were applied to minimize the geometry and remove steric clashes, followed by 10 ns of isothermal–isobaric (NPT) equilibration. The Berendsen barostat was applied to keep the pressure around 1 atm, while the temperature of 300 K was maintained throughout all the simulations with the V-rescale algorithm[48]. Molecular dynamics production runs (500-ns long) were carried out for all the systems in the canonical (NVT) ensemble, for a cumulative total of 2.5 μs. The particle mesh Ewald (PME)–Switch algorithm was used for electrostatic interactions with a cut-off of 1 nm, and a single cut-off of 1.2 nm was used for Van der Waals interactions. Four simulations for human GS and two for the GS of *Salmonella typhimurium* were run by placing the CoA moiety close to the adenosine binding site and allowing different initial positions for the palmitoyl tail. The CoA head invariably docked and remained tightly bound to the adenine binding site in all simulations. Among these, two favourable alternative arrangements (Extended Data Fig. 8b) for the tail were identified in both systems. In one of these conformations, the beginning of the palmitate tail (from the point of view of the CoA moiety) approaches very close to the conserved CYS209 (human residue numbering, conformation A in Extended Data Fig. 8b, details in Extended Data Fig. 8c), and in the other conformation

(conformation B in Extended Data Fig. 8b, details in Extended Data Fig. 8d) it approaches the conserved Ser65 and 75.

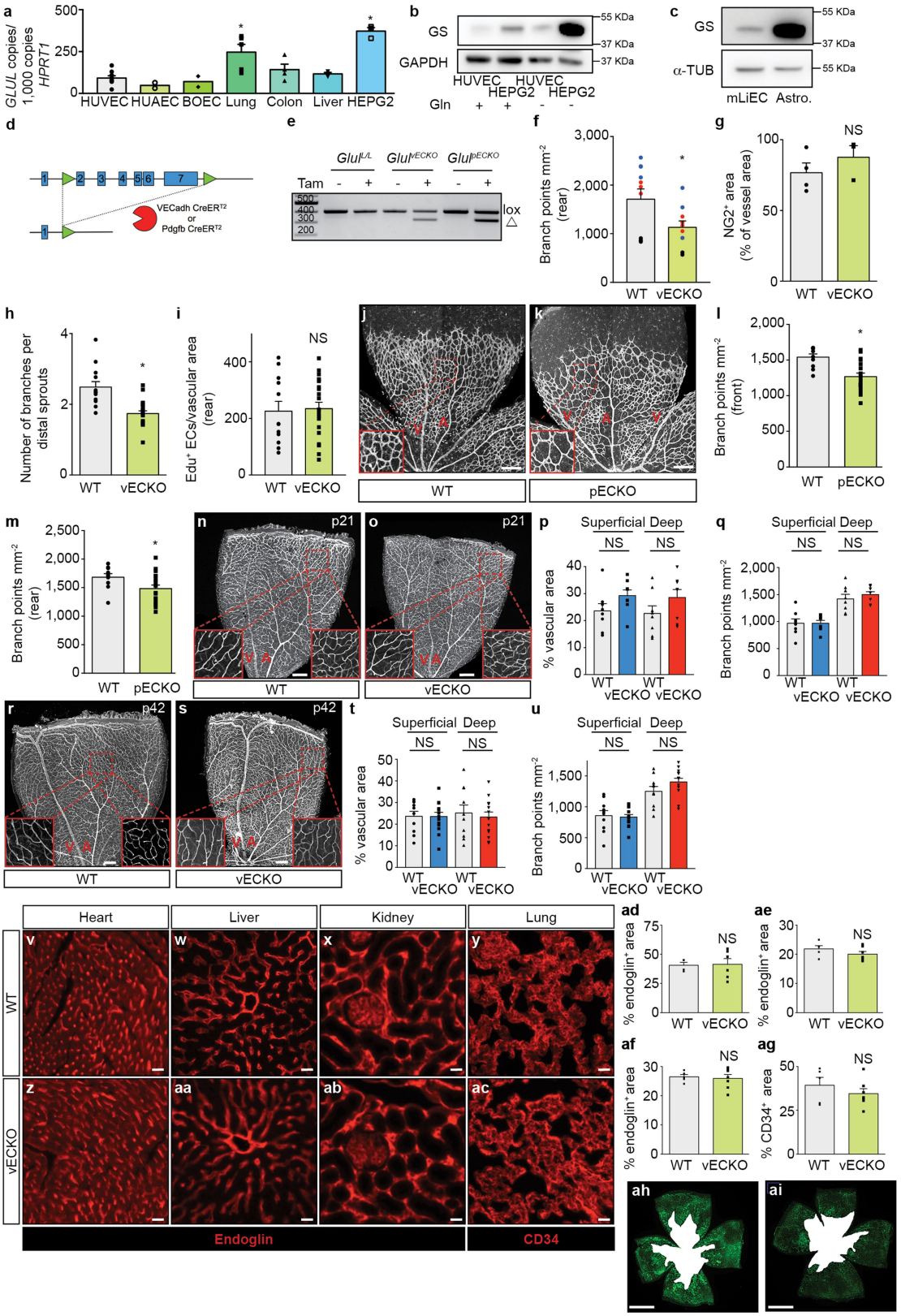
Multiple sequence alignments. A multiple sequence alignment of the GS protein across different species was performed with the Basic Local Alignment Search Tool (BLAST). The algorithm matches sequences according to local similarity, by optimizing their maximal segment pair score (MSP). The 100 matches with the highest identity to the *Homo sapiens* sequence surrounding amino acid C209 were taken from the UniProtKB/Swiss-Prot refined database.

Statistical analysis. Data represent mean ± s.e.m. of pooled experiments unless otherwise stated. Scatters in bar graphs represent the values of independent experiments or individual mice. In cases for which individual values are very similar, scatter points overlap and may no longer be visible as individual points. *n* values represent the number of independent experiments performed or the number of individual mice phenotyped. Statistical significance between groups was calculated with one of the following methods: for comparisons to point-normalized data, a two-tailed one-sample *t*-test was used in GraphPad Prism7. For pairwise comparisons, two-tailed unpaired *t*-tests were used in GraphPad Prism7. For multiple comparisons within one dataset, one-way ANOVA with Dunnett's multiple comparison (comparing every mean with the control mean rather than comparing every mean with every other mean) was used in GraphPad Prism7. Mixed-model statistics (this test does not assume normality or equal variance) was used with the experiment as a random factor only in cases for which confounding variation in baseline measurements between individual EC isolations (for each experiment, ECs were freshly isolated from individual human umbilical cords) or mouse litters precluded the use of the above statistical tests. For this, R and the lme4 package were used; *P* values were obtained with the Kenward–Roger *F*-test for small mixed-effect model datasets. In the most severe cases, the individual data points (each data point being the mean of the technical replicates within an experiment or an individual mouse) in the bar graphs have been colour-coded per experiment or per litter to show the baseline variation. The sample size for each experiment was not predetermined. A P value <0.05 was considered significant.

Reporting summary. Further information on research design is available in the Nature Research Reporting Summary linked to this paper.

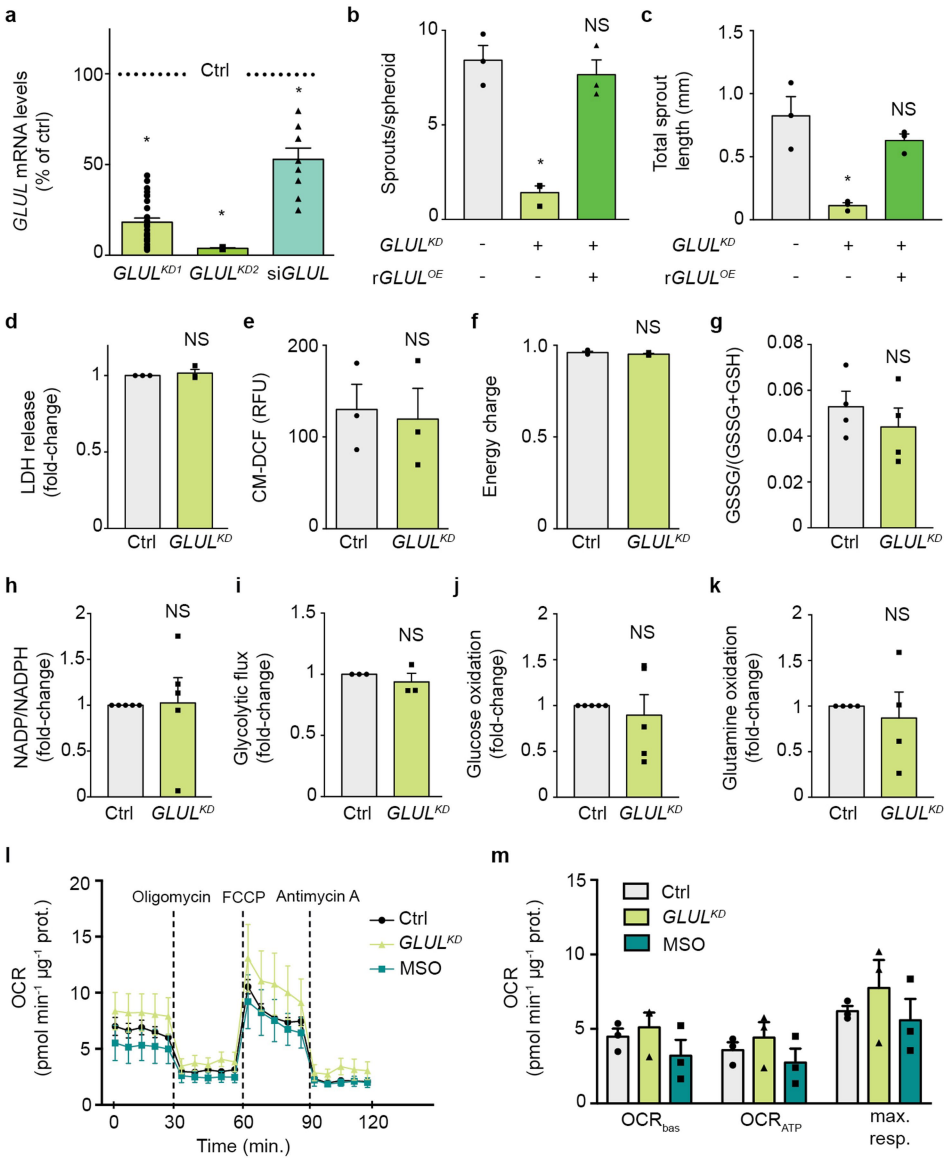
Data availability. Figures 1, 4, 5 and Extended Data Figs. 1, 7, and 8 have associated raw data (uncropped blots and/or gel pictures) in Supplementary Fig. 1. Figures 1, 2 and Extended Data Figs. 1, 4 have associated raw data (Excel files) for all bar graphs representing data from experiments involving mouse models. For the molecular modelling of palmitoyl-CoA docking into GS, models and trajectories are available on Figshare (doi: 10.6084/m9.figshare.6575438). Any additional information required to interpret, replicate or build upon the Methods or findings reported in the manuscript is available from the corresponding author upon request.

25. Zheng, B., Zhu, S. & Wu, X. Clickable analogue of cerulenin as chemical probe to explore protein palmitoylation. *ACS Chem. Biol*. **10**, 115–121 (2015).
26. Martin-Ramirez, J., Hofman, M., van den Biggelaar, M., Hebbel, R. P. & Voorberg, J. Establishment of outgrowth endothelial cells from peripheral blood. *Nat. Protoc*. **7**, 1709–1715 (2012).
27. Van Den Bosch, L., Vandenberghe, W., Klaassen, H., Van Houtte, E. & Robberecht, W. Ca^{2+}-permeable AMPA receptors and selective vulnerability of motor neurons. *J. Neurol. Sci*. **180**, 29–34 (2000).
28. McKinney, S. A., Murphy, C. S., Hazelwood, K. L., Davidson, M. W. & Looger, L. L. A bright and photostable photoconvertible fluorescent protein. *Nat. Methods* **6**, 131–133 (2009).
29. Mertens, N. M. A. C., Remaut, E. R. & Fiers, W. C. Regulatory system for inducible expression of genes with lambdoid promoters. Belgian patent WO 98/48025 (1998).
30. Mertens, N. M. A. C. & Kelly, A. G. Use of caspase enzymes for maturation of engineered recombinant polypeptide fusions. Belgian patent WO 04/074488 (2004).
31. Hamel, L. D., Deschenes, R. J. & Mitchell, D. A. A fluorescence-based assay to monitor autopalmitoylation of zDHHC proteins applicable to high-throughput screening. *Anal. Biochem*. **460**, 1–8 (2014).
32. Kümmel, D., Heinemann, U. & Veit, M. Unique self-palmitoylation activity of the transport protein particle component Bet3: a mechanism required for protein stability. *Proc. Natl Acad. Sci. USA* **103**, 12701–12706 (2006).
33. Fraccaroli, A. et al. Endothelial alpha-parvin controls integrity of developing vasculature and is required for maintenance of cell–cell junctions. *Circ. Res*. **117**, 29–40 (2015).
34. Ho, C. Y., Jaalouk, D. E., Vartiainen, M. K. & Lammerding, J. Lamin A/C and emerin regulate MKL1–SRF activity by modulating actin dynamics. *Nature* **497**, 507–511 (2013).
35. Timmerman, I. et al. A local VE-cadherin and Trio-based signaling complex stabilizes endothelial junctions through Rac1. *J. Cell Sci*. **128**, 3041–3054 (2015).
36. van Buul, J. D. et al. RHOG regulates endothelial apical cup assembly downstream from ICAM1 engagement and is involved in leukocyte trans-endothelial migration. *J. Cell Biol*. **178**, 1279–1293 (2007).
37. Heemskerk, N. et al. F-actin-rich contractile endothelial pores prevent vascular leakage during leukocyte diapedesis through local RHOA signalling. *Nat. Commun*. **7**, 10493 (2016).
38. Dedecker, P., Duwé, S., Neely, R. K. & Zhang, J. Localizer: fast, accurate, open-source, and modular software package for superresolution microscopy. *J. Biomed. Opt*. **17**, 126008 (2012).
39. Persson, F., Lindén, M., Unoson, C. & Elf, J. Extracting intracellular diffusive states and transition rates from single-molecule tracking data. *Nat. Methods* **10**, 265–269 (2013).
40. He, Y. et al. Glutamine synthetase deficiency in murine astrocytes results in neonatal death. *Glia* **58**, 741–754 (2010).
41. Benedito, R. et al. The notch ligands Dll4 and Jagged1 have opposing effects on angiogenesis. *Cell* **137**, 1124–1135 (2009).
42. Claxton, S. et al. Efficient, inducible Cre-recombinase activation in vascular endothelium. *Genesis* **46**, 74–80 (2008).
43. Pitulescu, M. E., Schmidt, I., Benedito, R. & Adams, R. H. Inducible gene targeting in the neonatal vasculature and analysis of retinal angiogenesis in mice. *Nat. Protoc*. **5**, 1518–1534 (2010).
44. Kenyon, B. M. et al. A model of angiogenesis in the mouse cornea. *Invest. Ophthalmol. Vis. Sci*. **37**, 1625–1632 (1996).
45. Abraham, M. J. et al. GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **1–2**, 19–25 (2015).
46. Maier, J. A. et al. ff14SB: improving the accuracy of protein side chain and backbone parameters from ff99SB. *J. Chem. Theory Comput*. **11**, 3696–3713 (2015).
47. Frisch, M. J. et al. *Gaussian 09 Rev*. C.01 (Gaussian, 2009).
48. Bussi, G., Donadio, D. & Parrinello, M. Canonical sampling through velocity rescaling. *J. Chem. Phys*. **126**, 014101 (2007).

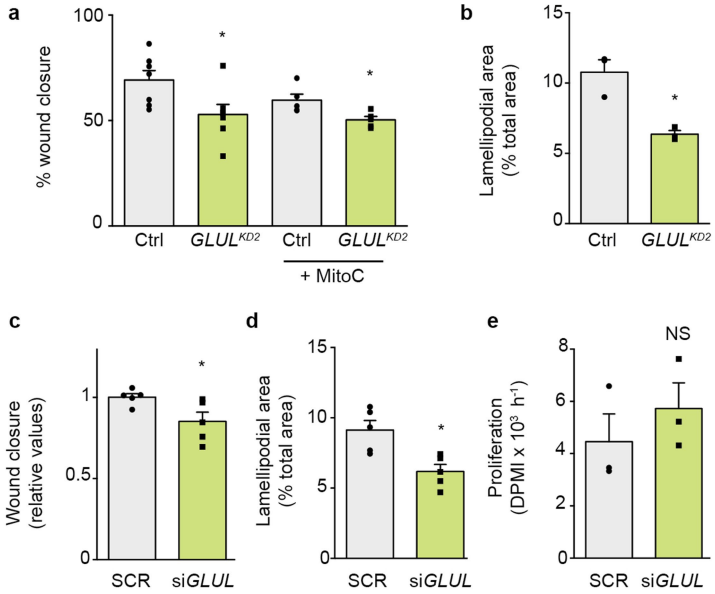


Extended Data Fig. 1 | See next page for caption.

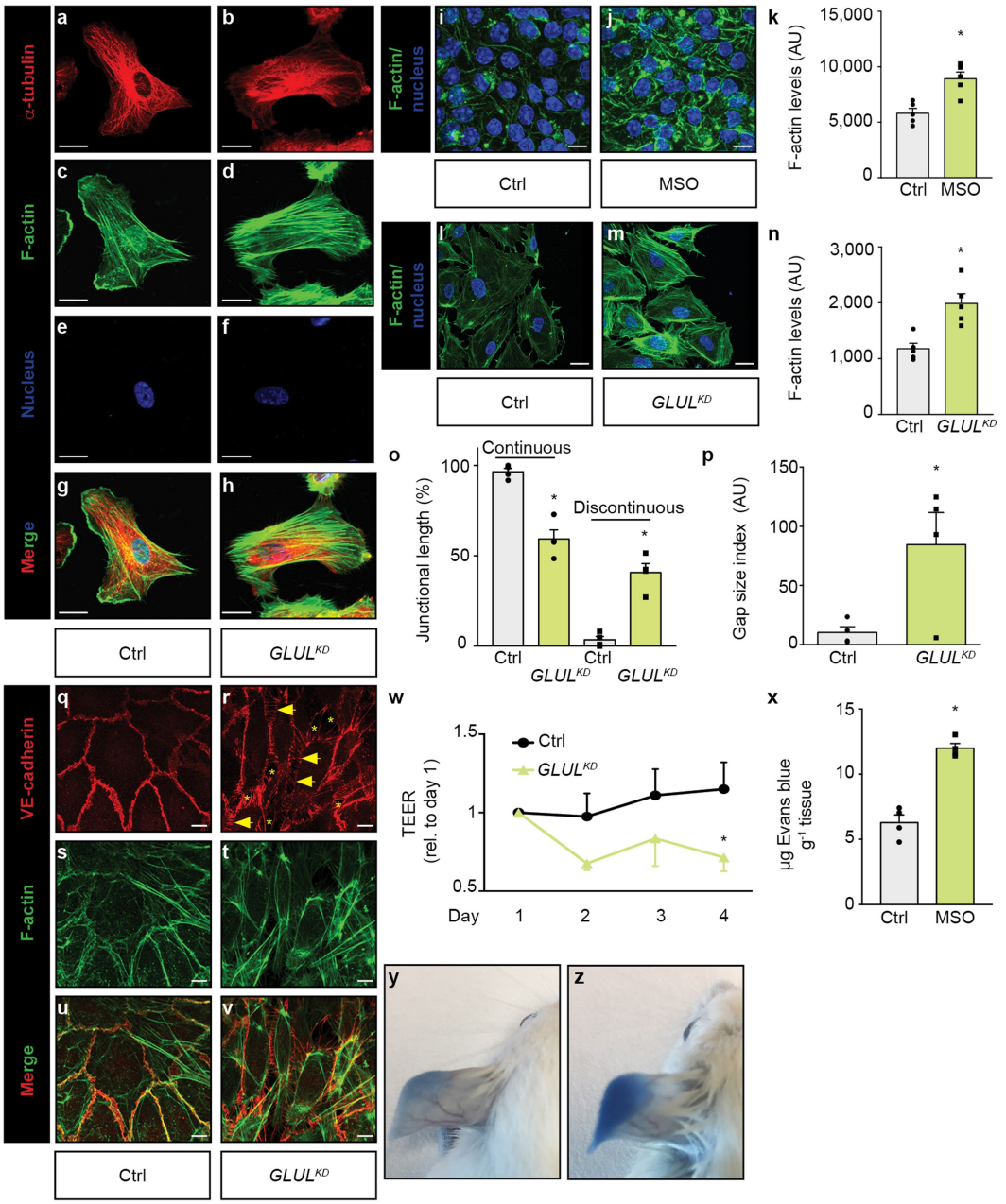
Extended Data Fig. 1 | *GLUL* knockout impairs vessel sprouting. **a**, *GLUL* mRNA levels in HUVECs ($n = 9$ donors), lung ECs ($n = 5$), colon ECs ($n = 4$), liver ECs ($n = 3$), human umbilical artery ECs (HUAECs) ($n = 2$) and human blood outgrowth ECs (BOECs) ($n = 2$); (mean ± s.e.m.; $^*P < 0.05$ versus HUVEC, Student's *t*-test) and in HEPG2 cells (mean ± s.e.m.; $n = 3$; $^*P < 0.05$ versus HUVEC, Student's *t*-test). **b**, **c**, Western blot of GS protein levels in HUVECs and HEPG2 cells in medium containing 0.6 mM glutamine (+) or 0.025 mM glutamine (−) (**b**), and in isolated mouse liver ECs (mLiECs) and mouse astrocytes (**c**) (representative immunoblots of two independent experiments are shown). **d**, **e**, Genomic organization of the *loxP*-flanked *Glul* allele before and after Cre-mediated excision (**d**) and correct recombination of the *lox* allele (L) in $Glul^{vECKO}$ and $Glul^{pECKO}$ mice upon tamoxifen (Tam) treatment, as assessed by genomic DNA PCR (**e**; the PCR to amplify the *loxP*-flanked *Glul* allele (lox) or to amplify the *cre*-recombined allele (Δ) were run in separate reactions but loaded in the same lane; the gel picture shown is representative of all control, $Glul^{vECKO}$ and $Glul^{pECKO}$ mice used in this study). **f**, Quantification of branchpoints at the rear of the plexus in $Glul^{vECKO}$ mice (mean ± s.e.m.; $n = 10$ mice for $Glul^{vECKO}$ and 11 for WT controls from 3 litters; $^*P < 0.05$ versus WT littermates, mixed-models R statistics). **g**, Pericyte coverage of retinal microvessels in WT and $Glul^{vECKO}$ littermates determined by NG2 staining and shown as the $NG2^+$ area as a percentage of the vessel area (mean ± s.e.m.; $n = 4$ mice for WT and 3 for $Glul^{vECKO}$ from 1 litter; NS, $P > 0.05$ versus WT, Student's *t*-test). **h**, Reduced complexity of the retinal vascular front in P5 $Glul^{vECKO}$ compared with WT mice, determined by the number of branches on distal sprouts (mean ± s.e.m.; $n = 13$ mice for WT and 21 for $Glul^{vECKO}$ from 5 litters; $^*P < 0.05$ versus WT, Student's *t*-test). **i**, Quantification of EdU^+ ECs at the rear of the plexus (mean ± s.e.m.; $n = 12$ mice for WT and 22 for $Glul^{vECKO}$ from 4 litters; NS, $P > 0.05$ versus WT littermates, Student's *t*-test). **j**–**m**, IB4 staining of P5 retinal vascular plexuses from WT (**j**) and $Glul^{pECKO}$ (**k**) mice (representative pictures with magnification shown in the inset; A, artery; V, vein) and quantification of branch points at the front (**l**) and the rear (**m**) of the plexus (mean ± s.e.m.; $n = 10$ mice for WT and 18 for $Glul^{pECKO}$ from 4 litters; $^*P \leq 0.05$ versus WT littermates, Student's *t*-test). **n**–**u**, IB4 staining of the retinal microvasculature of three-week-old (P21) (**n**, **o**) and six-week-old (P42) (**r**, **s**) WT and $Glul^{vECKO}$ littermates (A, artery; V, vein). The lower left insets display a higher magnification of the IB4-stained superficial plexus, whereas the lower right insets display a higher magnification of the deep plexus. The corresponding quantification of the vascular area (**p**, **t**) and the branch point density (**q**, **u**) in the superficial and the deep layer is also shown (mean ± s.e.m.; $n = 8$ mice for WT and 8 for $Glul^{vECKO}$ at P21, from two litters; $n = 10$ mice for WT and 14 for $Glul^{vECKO}$ at P42, from four litters; NS, $P > 0.05$ versus WT, Student's *t*-test). **v**–**ag**, Representative micrographs of heart (**v**, **z**), liver (**w**, **aa**) and kidney (**x**, **ab**) sections from WT and $Glul^{vECKO}$ littermates immunostained for the EC marker endoglin and of lung (**y**, **ac**) sections immunostained for the EC marker CD34 and corresponding quantifications of $endoglin^+$ (**ad**, heart; **ae**, liver; **af**, kidney) or $CD34^+$ (**ag**, lung) vascular area (mean ± s.e.m.; $n = 5$ mice (4 for heart) for WT and 7 (6 for heart) for $Glul^{vECKO}$, from two litters, NS, $P > 0.05$ versus WT, Student's *t*-test). **ah**–**ai**, Images of flat-mounted retinas from control (**ah**) and MSO-treated (**ai**) ROP mice (vaso-obliterated area in white). Images shown are representative of 7 (**ah**) and 6 (**ai**) mice. Exact *P* values: HUVEC versus lung ECs: 0.0278; HUVEC versus colon ECs: 0.1086; HUVEC versus liver ECs: 0.3334; HUVEC versus HEPG2: <0.0001 (**a**); <0.0001 (**f**); 0.3491 (**g**); <0.0001 (**h**); 0.8247 (**i**); 0.0012 (**l**); 0.050 (**m**); superficial: 0.1218; deep: 0.1720 (**p**); superficial: 0.9995; deep: 0.4289 (**q**); superficial: 0.9792; deep: 0.6602 (**t**); superficial: 0.7979; deep: 0.1275 (**u**); 0.9021 (**ad**); 0.2279 (**ae**); 0.7647 (**af**); 0.3614 (**ag**). Scale bars: 200 μm (**j**, **k**, **n**, **o**, **r**, **s**), 20 μm (**v**–**ac**), 1 mm (**ah**–**ai**). For gel source images, see Supplementary Fig. 1.



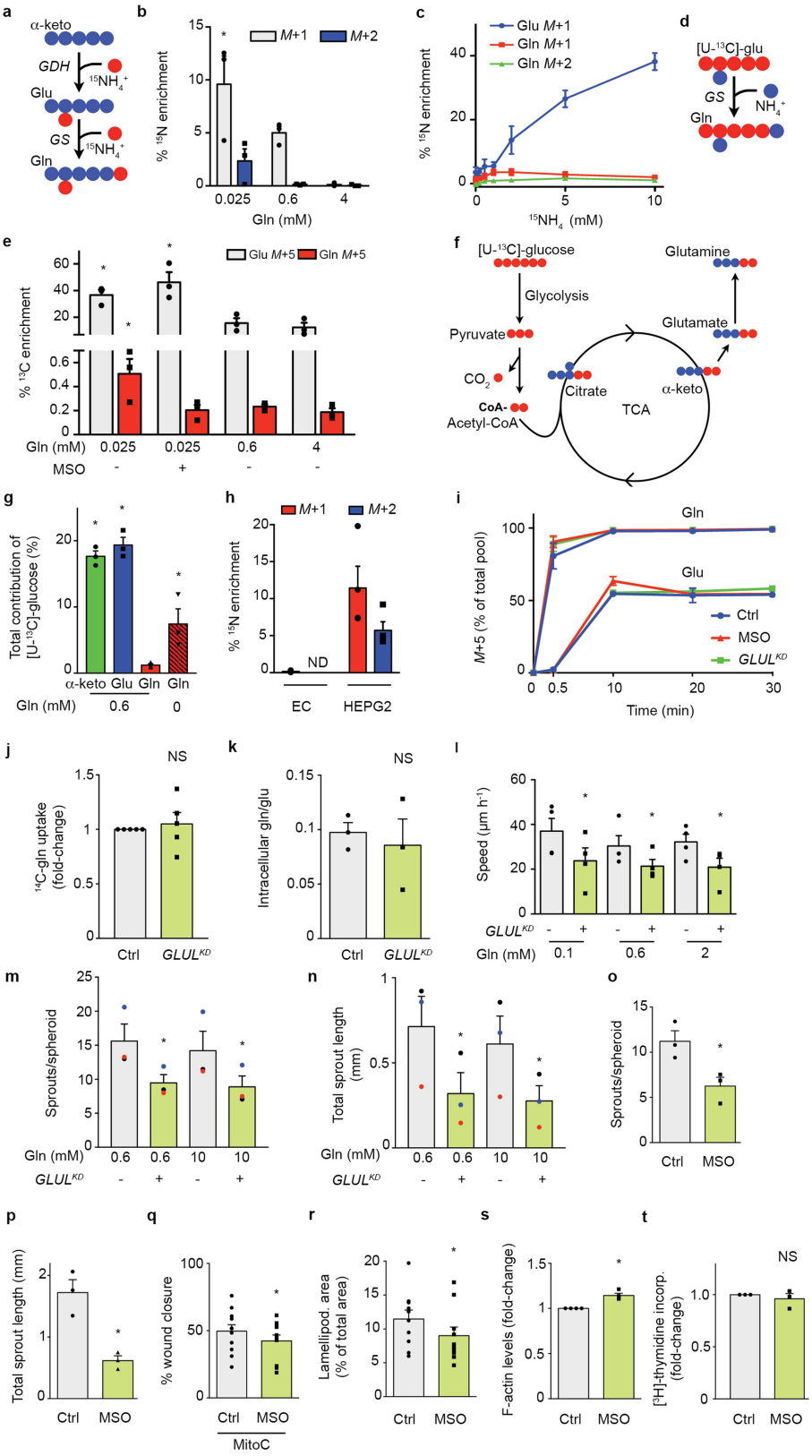
Extended Data Fig. 2 | Effects of silencing and pharmacological inhibition of GS on EC viability and central metabolism. a, *GLUL* mRNA levels in control ECs and ECs transduced with two different non-overlapping shRNAs targeting *GLUL* (*GLUL*^{KD1} and *GLUL*^{KD2}; *GLUL*^{KD1} is used in the experiments in the text and is denoted *GLUL*^{KD}) or transfected with scrambled siRNA (SCR) or siRNA targeting *GLUL* (si*GLUL*). Data are expressed as a percentage of the respective control, denoted by the horizontal dotted line (mean ± s.e.m.; $n = 28$ independent experiments for *GLUL*^{KD1}, $n = 3$ independent experiments for *GLUL*^{KD2} and $n = 9$ independent experiments for si*GLUL*; $^*P < 0.05$ versus the respective control; one-sample *t*-test). **b**, **c**, Quantification of number of sprouts (**b**) and total sprout length (**c**) for spheroid-sprouting assays with *GLUL*^{KD} ECs and *GLUL*^{KD} ECs expressing a shRNA-resistant *GLUL* mutant (r*GLUL*^{OE}) (mean ± s.e.m.; $n = 3$ independent experiments; $^*P < 0.05$ and NS, $P > 0.05$ versus control; one-way ANOVA with Dunnett's multiple comparison versus control). **d**, Viability of control (Ctrl) and *GLUL*^{KD} ECs as measured by lactate dehydrogenase (LDH) release assay (mean ± s.e.m.; $n = 3$ independent experiments; NS, $P > 0.05$ versus control, one-sample *t*-test). **e**, Intracellular levels of reactive oxygen species measured by CM-H_2DCFDA staining (mean ± s.e.m.; $n = 3$ independent experiments; NS, $P > 0.05$ versus control, Student's *t*-test). **f**, Energy charge measurement (([ATP] + 1/2[ADP]) / ([ATP] + [ADP] + [AMP])) in *GLUL*^{KD} and control ECs (mean ± s.e.m.; $n = 3$ independent experiments; NS, $P > 0.05$ versus control, Student's *t*-test). **g**, Ratio of oxidized glutathione (GSSG) over total glutathione levels (GSSG/(GSH + GSSG)) in *GLUL*^{KD} and control ECs (mean ± s.e.m.; $n = 4$ independent experiments; NS, $P > 0.05$ versus control, Student's *t*-test). **h**, NADP/NADPH ratio in *GLUL*^{KD} and control ECs (mean ± s.e.m.; $n = 5$ independent experiments; NS, $P > 0.05$ versus control, one-sample *t*-test). **i**–**k**, Effect of *GLUL* knockdown on major metabolic fluxes including glycolysis (**i**), glucose oxidation (**j**) and glutamine oxidation (**k**) (mean ± s.e.m.; $n = 3$ independent experiments for **i**, $n = 5$ for **j** and $n = 4$ for **k**; NS, $P > 0.05$ versus control, one-sample *t*-test). **l**, **m**, Oxygen consumption rate (OCR) in control, MSO-treated and *GLUL*^{KD} ECs in basal state and after injection of oligomycin, FCCP and antimycin A (**l**) (mean ± s.e.m.; $n = 3$ independent experiments), and calculation of OCR_{BAS}, OCR_{ATP} and maximal respiration (**m**) (mean ± s.e.m.; $n = 3$ independent experiments). Exact *P* values: *GLUL*^{KD1}: <0.0001; *GLUL*^{KD2}: <0.0001; si*GLUL*: <0.0001 (**a**); control versus *GLUL*^{KD}: 0.0147; control versus *GLUL*^{KD} + r*GLUL*^{OE}: 0.9824 (**b**); control versus *GLUL*^{KD}: 0.0083; control versus *GLUL*^{KD} + r*GLUL*^{OE}: 0.6528 (**c**); 0.5717 (**d**); 0.8206 (**e**); 0.3715 (**f**); 0.4398 (**g**); 0.9291 (**h**); 0.4691 (**i**); 0.6643 (**j**); 0.6786 (**k**). CM-DCF, CM-H_2DCFDA; OCR_{BAS}, basal oxygen consumption rate; OCR_{ATP}, ATP-generating oxygen consumption rate; RFU, relative fluorescence units.



Extended Data Fig. 3 | *GLUL* knockdown reduces EC motility. a, Wound closure in control and $GLUL^{KD2}$ EC monolayer scratch assays with or without pretreatment with mitomycin C (mean ± s.e.m.; $n = 7$ and 5 independent experiments with and without mitomycin C, respectively; $^*P < 0.05$ versus corresponding control; Student's *t*-test). **b**, Quantification of lamellipodial area (as a percentage of total cellular area) in control and $GLUL^{KD2}$ ECs (mean ± s.e.m.; $n = 3$ independent experiments; $^*P < 0.05$ versus control; Student's *t*-test). **c**, Wound closure in monolayer scratch assays with SCR- and si*GLUL*-transfected ECs (mean ± s.e.m.; $n = 5$ independent experiments; $^*P < 0.05$ versus SCR; Student's *t*-test). **d**, Quantification of lamellipodial area (as a percentage of total cellular area) in SCR- and si*GLUL*-transfected ECs (mean ± s.e.m.; $n = 5$ independent experiments; $^*P < 0.05$ versus SCR; Student's *t*-test). **e**, Proliferation of SCR- and si*GLUL*-transfected ECs, as measured by $[^3H]$ thymidine incorporation into DNA (mean ± s.e.m.; $n = 3$ independent experiments; NS, $P > 0.05$ versus SCR; Student's *t*-test). Exact *P* values: control versus $GLUL^{KD2}$: 0.0290; control versus $GLUL^{KD2}$ + MitoC: 0.0223 (**a**); 0.0088 (**b**); 0.0407 (**c**); 0.0083 (**d**); 0.4335 (**e**).

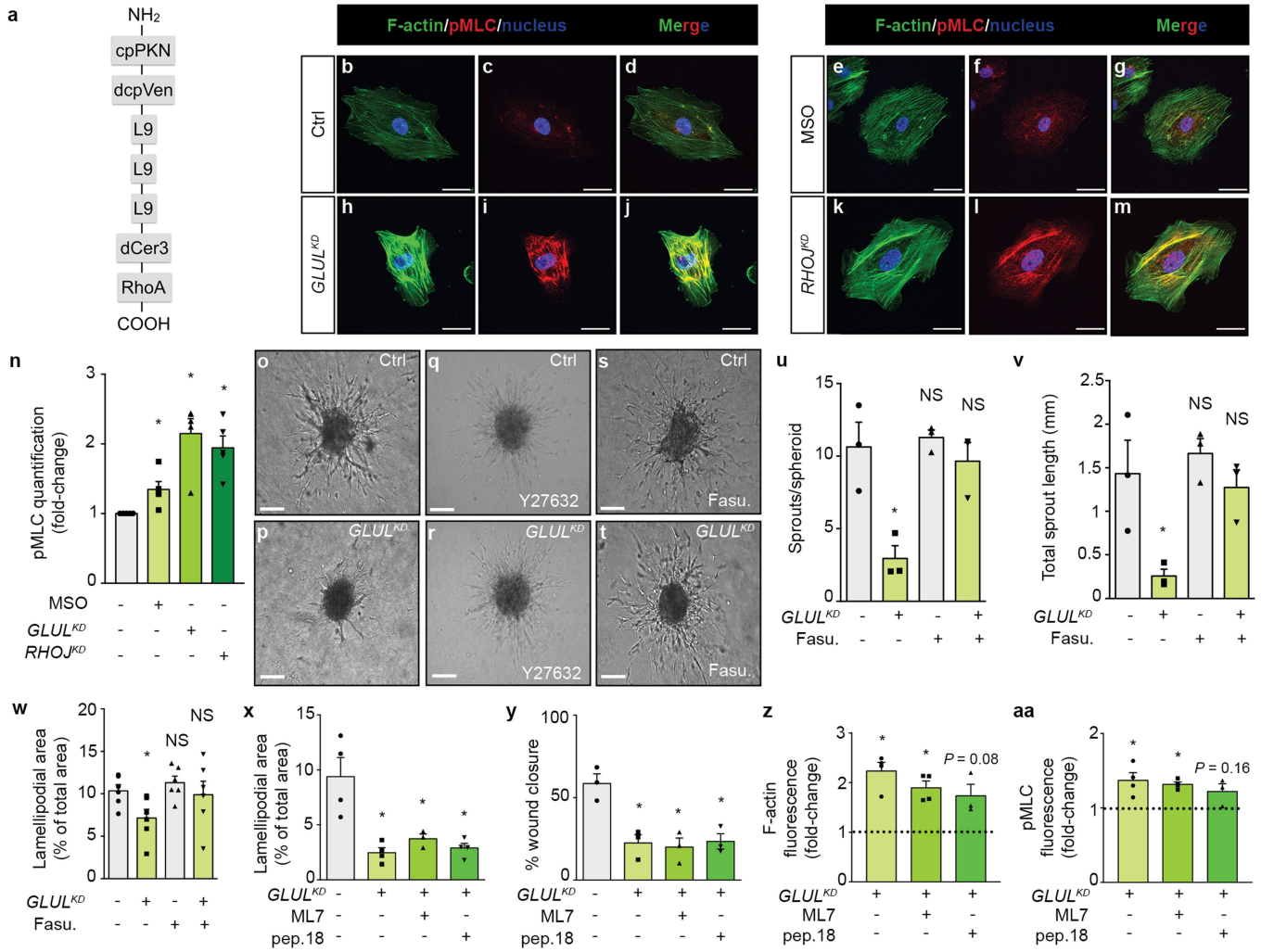


Extended Data Fig. 4 | Effects of *GLUL* silencing on cytoskeleton and barrier function. a–h, Images of control (**a**, **c**, **e**, **g**) and $GLUL^{KD}$ (**b**, **d**, **f**, **h**) ECs after staining for α-tubulin (**a**, **b**), F-actin (**c**, **d**) and nuclear staining (**e**, **f**); **g** and **h** show merged images. The images shown are representative of 3 independent experiments. **i–k**, Representative images of phalloidin (F-actin) + Hoechst-stained liver ECs 6 h after isolation from control (**i**) and MSO-treated (**j**) mice, and corresponding quantification of F-actin levels (**k**) (mean ± s.e.m.; $n = 5$ mice per group; $^*P < 0.05$ versus control, Student's *t*-test). **l–n**, Representative images of phalloidin-stained confluent monolayer control (**l**) and $GLUL^{KD}$ (**m**) ECs aligning a scratch wound, and quantification of F-actin levels (**n**) (mean ± s.e.m.; $n = 5$ independent experiments; $^*P < 0.05$ versus control, Student's *t*-test). **o**, Quantification of the length of discontinuous and continuous VE-cadherin-stained junctions in control and $GLUL^{KD}$ ECs (mean ± s.e.m.; $n = 4$ independent experiments; $^*P < 0.05$ versus control, Student's *t*-test). **p**, Quantification of VE-cadherin gap size index in control and $GLUL^{KD}$ EC monolayers (mean ± s.e.m.; $n = 4$ independent experiments; $^*P < 0.05$ versus control, Student's *t*-test). **q–v**, Corresponding representative images of monolayer control (**q**, **s**, **u**) and $GLUL^{KD}$ (**r**, **t**, **v**) ECs stained for VE-cadherin (**q**, **r**, **u**, **v**) and F-actin (**s**, **t**, **u**, **v**). Yellow arrows in **r** point to discontinuous VE-cadherin junctions and yellow asterisks indicate intracellular gaps. **w**, Quantification of transendothelial electrical resistance in control and $GLUL^{KD}$ EC monolayers (mean ± s.e.m.; $n = 4$ independent experiments; $^*P < 0.05$ versus control, Student's *t*-test at each time point). **x–z**, Quantification (**x**) of Evans blue dye extracted from the ears of control and MSO-treated mice, induced by topical application of mustard oil ($n = 4$ mice for each condition, $^*P < 0.05$; Student's *t*-test), and representative pictures of the leakage of Evans blue dye into the ear tissue in control (**y**) and MSO-treated (**z**) mice. Exact *P* values: 0.0030 (**k**); 0.0036 (**n**); continuous control versus $GLUL^{KD}$: 0.0005; discontinuous control versus $GLUL^{KD}$: 0.0005 (**o**); 0.0356 (**p**); 0.0181 (**w**); 0.0002 (**x**). Scale bars: 20 µm (**a–h**, **l**, **m**), 10 µm (**i**, **j**, **q–v**). AU, arbitrary units.

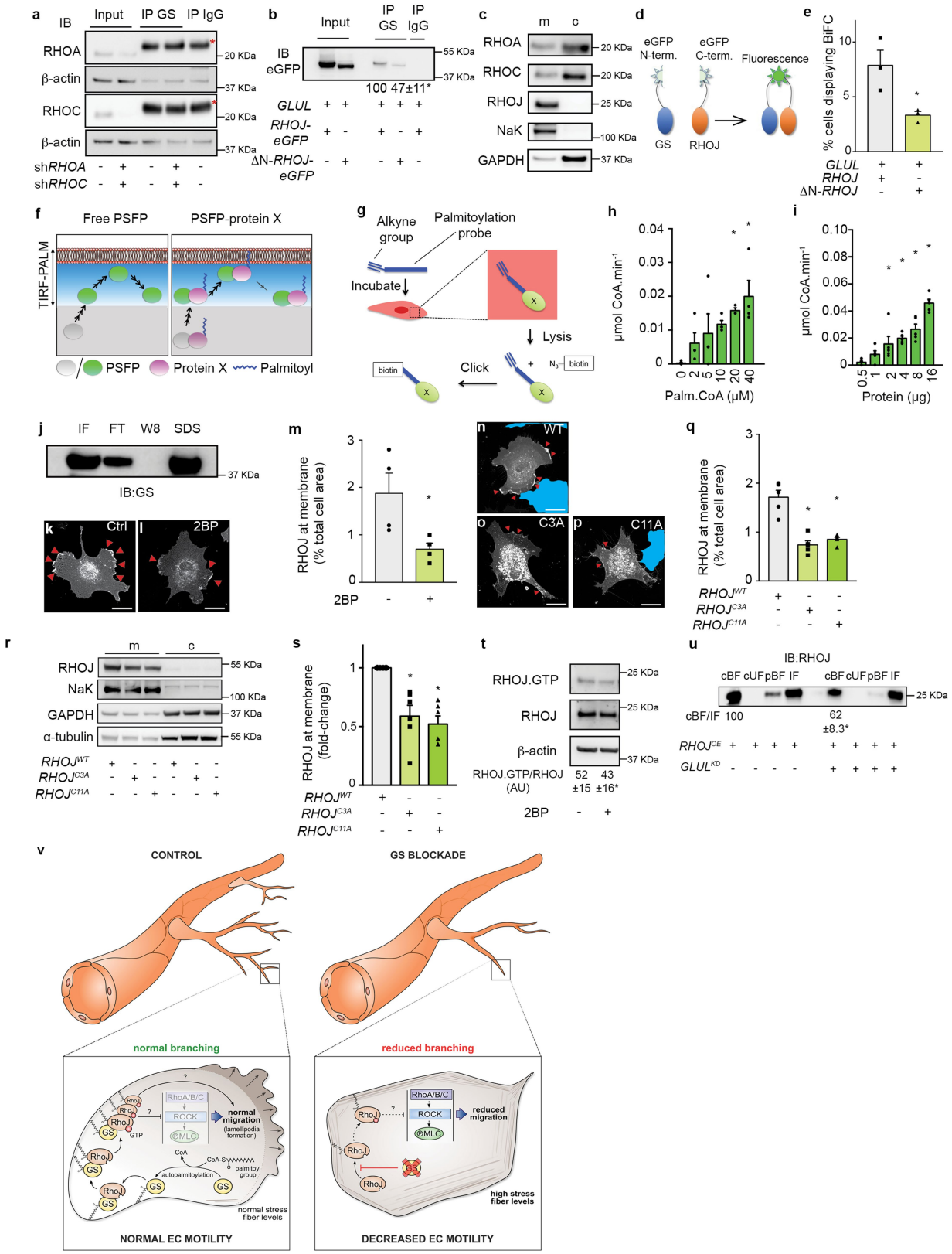


Extended Data Fig. 5 | See next page for caption.

Extended Data Fig. 5 | Enzymatic activity of GS and its role in EC migration. a, Scheme for the $^{15}NH_4^+$ labelling of glutamate and glutamine with unlabelled carbons (blue) and labelled nitrogens (red). **b**, ^{15}N incorporation into glutamine (measured as the percentage of isotope enrichment in glutamine either as $M+1$ (singly labelled) or $M+2$ (doubly labelled), 30 min after adding $^{15}NH_4^+$) in medium with dialysed serum and different glutamine concentrations (mean ± s.e.m.; $n=3$ independent experiments; one-way ANOVA with Dunnett's multiple comparisons versus 4 mM; $^*P<0.05$). **c**, ^{15}N incorporation into glutamate (measured as the percentage of isotope enrichment in $M+1$) and glutamine (measured as the percentage of isotope enrichment in $M+1$ (singly labelled) and $M+2$ (doubly labelled)), 30 min after adding increasing concentrations of $^{15}NH_4Cl$ (mean ± s.e.m.; $n=3$ independent experiments). **d**, Scheme of glutamine labelling from [U-^{13}C] glutamate with unlabelled nitrogens (blue) and labelled carbons (red). **e**, Contribution of labelled [U-^{13}C]glutamate to intracellular glutamine at various glutamine concentrations (percentage of isotope enrichment in $M+5$ glutamine and glutamate, 30 min after adding the tracer) (mean ± s.e.m.; $n=3$ independent experiments; one-way ANOVA with Dunnett's multiple comparisons versus 4 mM; $^*P<0.05$). **f**, Scheme for the contribution of carbons from [U-^{13}C]glucose to glutamine, with labelled carbons (red) and unlabelled carbons (blue). Incorporation is shown after one turn of the tricarboxylic acid (TCA) cycle. **g**, Total contribution of carbons from [U-^{13}C]glucose to α-ketoglutarate, glutamate and glutamine in ECs in medium with or without glutamine, 48 h after adding the tracer (mean ± s.e.m.; $n=3$ independent experiments; $^*P<0.05$ versus total contribution in glutamine at 0.6 mM external glutamine, one-way ANOVA with Dunnett's multiple comparisons). **h**, Incorporation of ^{15}N into glutamine (measured as the percentage of isotope enrichment in $M+1$ (singly labelled) and $M+2$ (doubly labelled), 30 min after adding $^{15}NH_4^+$) in ECs and HEPG2 cells (mean ± s.e.m.; $n=4$ independent experiments. ND, not detected). **i**, ^{13}C-glutamine uptake kinetics in control, MSO-treated and $GLUL^{KD}$ ECs and subsequent conversion to glutamate. See Methods for explanation of the different time points. Data are for the $M+5$ isotopomer, as a percentage of the total intracellular glutamine or glutamate pool (mean ± s.e.m.; $n=3$ independent experiments, except for 30 min for which $n=1$ experiment; no statistical differences between control, MSO-treated and $GLUL^{KD}$ ECs were observed for glutamine or for glutamate; one-way ANOVA with Dunnett's multiple comparison versus control at each time point; no statistical analysis was performed at 30 min). **j**, ^{14}C-glutamine uptake in control and $GLUL^{KD}$ ECs (mean ± s.e.m.; $n=5$ independent experiments; NS, $P>0.05$ versus control, one-sample t-test). **k**, Ratio of intracellular glutamine and glutamate levels in control and $GLUL^{KD}$ ECs (mean ± s.e.m.; $n=3$ independent experiments; NS, $P>0.05$ versus control, Student's t-test). **l**, Velocity measurement of control and $GLUL^{KD}$ ECs at different glutamine concentrations (mean ± s.e.m.; $n=4$ independent experiments; $^*P<0.05$ versus corresponding control, mixed-models R statistics). **m**, **n**, Effect of glutamine concentration on sprout number (**m**) and total sprout length (**n**) in control and $GLUL^{KD}$ spheroids (mean ± s.e.m.; $n=3$ independent experiments; $^*P<0.05$ versus corresponding control, mixed-models R statistics). **o**, **p**, Number of sprouts per spheroid (**o**) and total sprout length (**p**) in control and MSO-treated EC spheroids (mean ± s.e.m.; $n=3$ independent experiments; $^*P<0.05$ versus control, paired Student's t-test). **q**–**s**, Effect of MSO-treatment on EC motility parameters: wound closure of mitomycin C-treated ECs (**q**) (mean ± s.e.m.; $n=11$ independent experiments; $^*P<0.05$ versus control, Student's t-test), lamellipodial area (**r**) (mean ± s.e.m.; $n=10$ independent experiments; $^*P<0.05$ versus control, paired Student's t-test) and F-actin levels, 1 h after latrunculin wash-out (**s**) (mean ± s.e.m.; $n=4$ independent experiments; $^*P<0.05$ versus control, one-sample t-test). **t**, [^3H]Thymidine incorporation in control and MSO-treated ECs (mean ± s.e.m.; $n=3$ independent experiments; NS, $P>0.05$ versus control, one-sample t-test). Exact P values: $M+1$ 0.025 mM versus $M+1$ 4 mM: 0.0096; $M+1$ 0.6 mM versus $M+1$ 4 mM: 0.1206; $M+2$ 0.025 mM versus $M+2$ 4 mM: 0.0839; $M+2$ 0.6 mM versus $M+2$ 4 mM: 0.9921 (**b**); Glu $M+5$ 0.6 mM versus Glu $M+5$ 4 mM: 0.9372; Glu $M+5$ 0.025 mM + MSO versus Glu $M+5$ 4 mM: 0.0034; Glu $M+5$ 0.025 mM versus Glu $M+5$ 4 mM: 0.0215; Gln $M+5$ 0.6 mM versus Gln $M+5$ 4 mM: 0.9297; Gln $M+5$ 0.025 mM + MSO versus Gln $M+5$ 4 mM: 0.9961; Gln $M+5$ 0.025 mM versus Gln $M+5$ 4 mM: 0.0268 (**e**); α-keto 0.6 mM versus Gln 0.6 mM: 0.0001; Glu 0.6 mM versus Gln 0.6 mM: 0.0001; Gln 0 mM versus Gln 0.6 mM: 0.0285 (**g**); Gln 0.5 min: control versus MSO: 0.4846; control versus $GLUL^{KD}$: 0.5904; Gln 10 min: control versus MSO: 0.6709; control versus $GLUL^{KD}$: 0.6910; Gln 20 min: control versus MSO: 0.5896; control versus $GLUL^{KD}$: 0.6784; Glu 0.5 min: control versus MSO: 0.9774; control versus $GLUL^{KD}$: 0.8810; Glu 10 min: control versus MSO: 0.0502; control versus $GLUL^{KD}$: 0.9598; Glu 20 min: control versus MSO: 0.9782; control versus $GLUL^{KD}$: 0.7783 (**i**); 0.6623 (**j**); 0.6704 (**k**); control versus $GLUL^{KD}$ 0.1 mM: 0.0054; control versus $GLUL^{KD}$ 0.6 mM: 0.0247 control versus $GLUL^{KD}$ 2 mM: 0.0017 (**l**); control versus $GLUL^{KD}$ 0.6 mM and 10 mM: <0.0001 (**m**); control versus $GLUL^{KD}$ 0.6 mM and 10 mM: <0.0001 (**n**); 0.0313 (**o**); 0.0075 (**p**); 0.0019 (**q**); 0.0116 (**r**); 0.0091 (**s**); 0.5110 (**t**). α-keto, α-ketoglutarate; GDH, glutamate dehydrogenase.

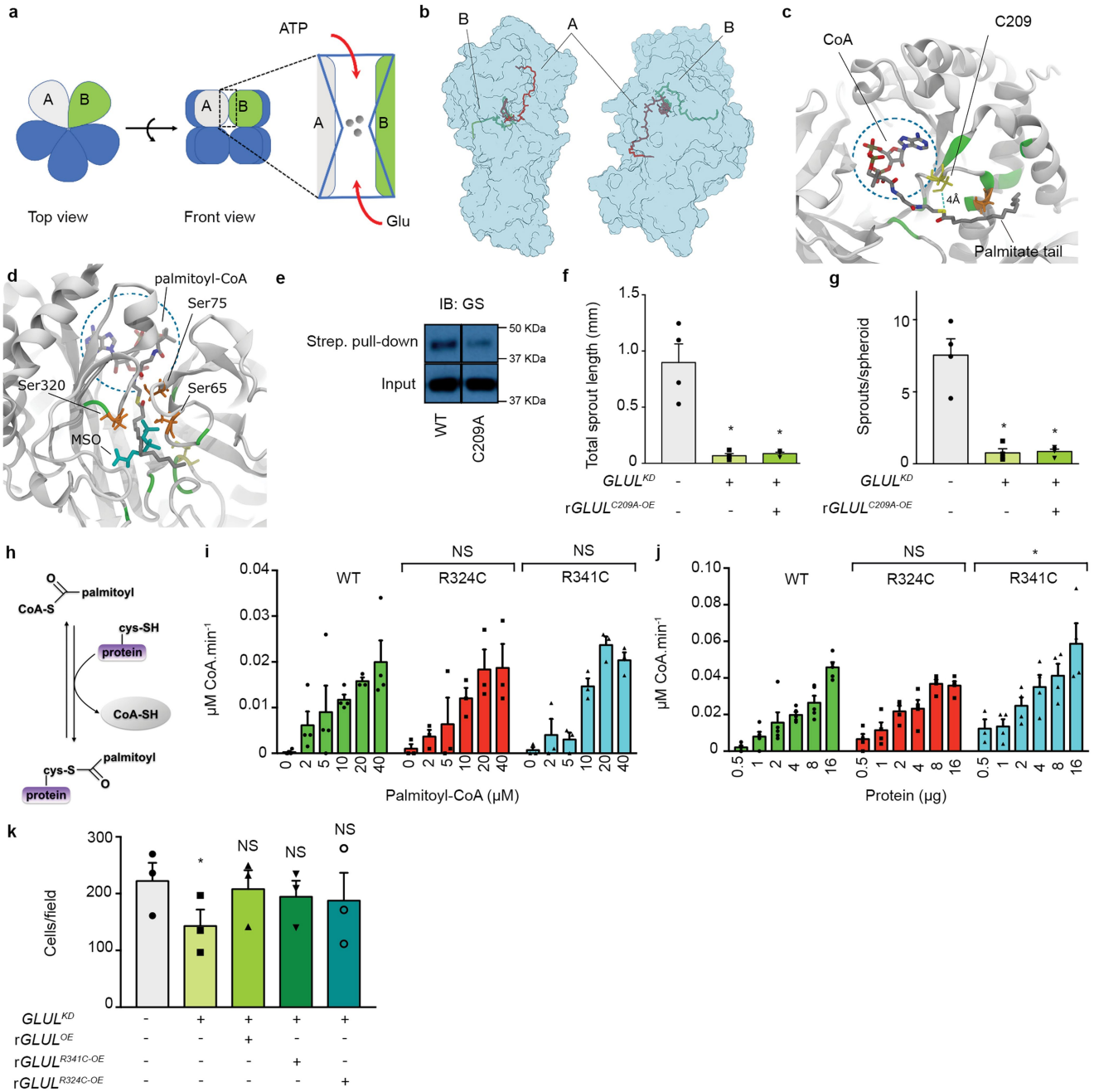


Extended Data Fig. 6 | Rescuing the phenotype associated with *GLUL*^{KD} in vitro. a, Schematic of the DORA–RHOA–FRET biosensor, depicting from N- to C-terminal the circular permutated RHOA effector protein kinase N (cpPKN), the dimeric circular permutated Venus (dcpVen), the ribosomal protein-based linkers (L9), the dimeric Cerulean3 (dCer3) and RHOA. **b**–**m**, Representative images of control (**b**–**d**), MSO-treated (**e**–**g**), *GLUL*^{KD} (**h**–**j**) and *RHOJ*^{KD} (**k**–**m**) ECs after staining for F-actin (phalloidin) (**b**, **d**, **e**, **g**, **h**, **j**, **k**, **m**) and pMLC (**c**, **d**, **f**, **g**, **i**, **j**, **l**, **m**). **n**, Quantification of the pMLC immunoreactivity (mean ± s.e.m.; $n = 5$ independent experiments; $^*P < 0.05$ versus control, one-sample *t*-test). **o**–**t**, Representative images of control (**o**, **q**, **s**) and *GLUL*^{KD} (**p**, **r**, **t**) EC spheroids treated with vehicle (**o**, **p**) or the ROCK inhibitors Y27632 (**q**, **r**) or fasudil hydrochloride (Fasu.; **s**, **t**). **u**, **v**, Quantification of the number of sprouts per spheroid (**u**) and sprout length (**v**) (mean ± s.e.m.; $n = 3$ independent experiments; $^*P < 0.05$ and NS, $P > 0.05$ versus untreated control, one-way ANOVA with Dunnett's multiple comparisons versus untreated control). **w**, Quantification of the lamellipodial area in vehicle- or fasudil-hydrochloride-treated control and *GLUL*^{KD} ECs (mean ± s.e.m.; $n = 6$ independent experiments; $^*P < 0.05$ and NS, $P > 0.05$ versus untreated control, one-way ANOVA with Dunnett's multiple comparisons versus untreated control). **x**, Quantification of the lamellipodial area in vehicle-, ML7- or peptide-18 (pep.18)-treated *GLUL*^{KD} and control ECs (mean ± s.e.m.; $n = 4$ independent experiments of which 3 experiments included the ML7-treatment; $^*P < 0.05$ versus untreated control, one-way ANOVA with Dunnett's multiple comparisons versus untreated control). **y**, Scratch wound closure in vehicle-, ML7- or peptide-18-treated *GLUL*^{KD} and control ECs (mean ± s.e.m.; $n = 3$ independent experiments; $^*P < 0.05$ versus untreated control, one-way ANOVA with Dunnett's multiple comparisons versus untreated control). **z**, Fold changes (versus untreated control ECs) in F-actin levels from phalloidin-stained vehicle-, ML7- or peptide-18-treated *GLUL*^{KD} ECs (mean ± s.e.m.; $n = 4$ independent experiments of which 3 included the peptide 18-treatment; $^*P < 0.05$ versus untreated control, one-sample *t*-test). **aa**, Fold changes (versus untreated control ECs) in pMLC levels from pMLC-immunostained vehicle-, ML7- or peptide-18-treated *GLUL*^{KD} ECs (mean ± s.e.m.; $n = 4$ independent experiments of which 3 included the peptide 18-treatment; $^*P < 0.05$ versus untreated control, one-sample *t*-test. Exact *P* values: MSO: 0.0372; *GLUL*^{KD}: 0.0060; *RHOJ*^{KD}: 0.0051 (**n**); *GLUL*^{KD} versus control: 0.0045; Fasu. versus control: 0.9596; *GLUL*^{KD} + Fasu. versus control: 0.8857 (**u**); *GLUL*^{KD} versus control: 0.0199; Fasu. versus control: 0.8309; *GLUL*^{KD} + Fasu. versus control: 0.9327 (**v**) *GLUL*^{KD} versus control: 0.0074; Fasu. versus control: 0.5906; *GLUL*^{KD} + Fasu. versus control: 0.9900; (**w**); *GLUL*^{KD} versus control: 0.0011; *GLUL*^{KD} + ML7 versus control: 0.0079; *GLUL*^{KD} + pep.18 versus control: 0.0017 (**x**); *GLUL*^{KD} versus control: 0.0034; *GLUL*^{KD} + ML7 versus control: 0.0022; *GLUL*^{KD} + pep.18 versus control: 0.0040 (**y**); *GLUL*^{KD}: 0.0058; ML7: 0.0072; pep.18: 0.0888 (**z**); *GLUL*^{KD}: 0.0369; ML7: 0.0021; pep.18: 0.1672 (**aa**). Scale bars: 20 μm (**b**–**m**), 100 μm (**o**–**t**). For gel source images, see Supplementary Fig. 1.



Extended Data Fig. 7 | See next page for caption.

Extended Data Fig. 7 | Rho GTPase localization and interaction with GS. a, Co-IP assays showing no detectable interaction between GS and RHOA or RHOC (red asterisk indicates a non-specific band (also present in the IgG controls and unaffected by silencing of *RHOA* or *RHOC*)). Image shown is representative of 3 independent experiments. **b**, Co-IP of overexpressed *GLUL* and *RHOJ-eGFP* or *ΔN20-RHOJ-eGFP* in ECs. Quantifications are mean ± s.e.m.; $n = 4$ independent experiments; $^*P < 0.05$, one-sample *t*-test. In some of the experiments, the expression of ΔN20-RHOJ–eGFP was lower than the expression of RHOJ–eGFP. To correct for this, densitometric quantification was performed and signals in immunoprecipitation lanes were normalized to input signals. **c**, Immunoblotting for RHOA and RHOC on cytosolic (c) and membrane (m) fractions of ECs with NaK as membrane marker and GAPDH as cytosolic marker. Image shown is representative of 3 independent experiments. **d**, BiFC assay with GS coupled to the N-terminal half of eGFP, and RHOJ coupled to the C-terminal half of eGFP. Only when GS and RHOJ are in close proximity do the two eGFP half-sites complement each other and form a functional eGFP. **e**, Percentage of ECs displaying BiFC upon overexpression of *GLUL-eGFP*$^{1/2}$ and *RHOJ-eGFP*$^{2/2}$ or *GLUL-eGFP*$^{1/2}$ and *ΔN20-RHOJ-eGFP*$^{2/2}$. Data are mean ± s.e.m.; $n = 3$ independent experiments; $^*P < 0.05$; Student's *t*-test. **f**, Schematic of SPT-PALM imaging under TIRF illumination with the plasma membrane depicted at the top. The TIRF region is bright (whereas the part outside the TIRF region is greyed out) and contains the plasma membrane and its immediately adjacent space (not shown at exact relative dimensions). Weight and number of arrowheads represent the velocity of single particles (the photoswitchable fluorescent protein (PSFP) or the PSFP coupled to the protein of interest (here GS)). The PSFP is activated upon entry into the TIRF region and is colour-coded differently inside and outside of the TIRF region. PSFP–GS displays reduced velocity in the TIRF region, presumably because of palmitoylation and membrane association of GS. **g**, Scheme for the in-cell labelling of proteins with clickable alkyne-containing palmitoylation probes and subsequent biotin-azide clicking. X represents a palmitoylated protein, N_3 is the biotin-coupled azide. **h**, **i**, Rate of CoA release from palmitoyl-CoA as a readout for recombinant human GS autopalmitoylation while varying either the doses of palmitoyl-CoA (**h**) or the amounts of recombinant GS (**i**) (mean ± s.e.m.; $n = 4$ independent experiments for **h** and $n = 5$ for **i**; $^*P < 0.05$, one-way ANOVA with Dunnett's multiple comparisons versus 0 μM palmitoyl-CoA or versus 0.5 μg recombinant GS). **j**, Representative GS immunoblot (of 3 independent experiments) for the binding of recombinant human GS to palmitoyl-CoA agarose. IF, input fraction; FT, flow through; W8, wash fraction 8; SDS is the eluate. **k**–**m**, Representative images of RHOJ–eGFP localization in ECs under vehicle-treatment (**k**) or treatment with 2BP (a pan-palmitoylation inhibitor) (**l**). Red arrowheads indicate eGFP signal at membrane ruffles, which was quantified as the percentage of the total cellular area (**m**) (mean ± s.e.m.; $n = 4$ independent experiments; $^*P < 0.05$ versus vehicle-treated, paired Student's *t*-test). **n**–**p**, Representative images of ECs overexpressing wild-type *RHOJ-eGFP* (**n**), *RHOJ-eGFP* encoding the point mutation C3A (**o**) or *RHOJ-eGFP* encoding the point mutation C11A (**p**). Red arrowheads indicate RHOJ at the plasma membrane. ECs that are not completely in the field of view have been masked out in blue. **q**, Quantification of the RHOJ–eGFP positive area at the plasma membrane as a percentage of the total cell area. Data are mean ± s.e.m.; $n = 5$ independent experiments; $^*P < 0.05$; one-way ANOVA with Dunnett's comparison versus wild-type RHOJ. **r**, RHOJ immunoblotting on membrane versus cytosolic fractions from ECs overexpressing wild-type *RHOJ-eGFP* (*RHOJ*^{WT}), *RHOJ-eGFP* encoding the point mutation C3A (*RHOJ*^{C3A}) or *RHOJ-eGFP* encoding the point mutation C11A (*RHOJ*^{C11A}), with NaK as membrane marker and GAPDH and α-tubulin as cytosolic markers. **s**, Densitometric quantification of RHOJ/NaK as determined in **r**. Data are mean ± s.e.m.; $n = 6$ independent experiments; $^*P < 0.05$; one-sample *t*-test. **t**, RHOJ activity in ECs under treatment with vehicle or 2BP (blots are representative of 3 independent experiments; densitometric quantification in arbitrary units is mean ± s.e.m.; $^*P < 0.05$, paired Student's *t*-test versus vehicle-treated). **u**, RHOJ immunoblotting of control and *GLUL*^{KD} ECs overexpressing RHOJ (RHOJ^{OE}) subjected to acyl-resin-assisted capture. The cleaved bound fraction (cBF) represents palmitoylated RHOJ. IF is the input fraction, whereas the cleaved unbound fraction (cUF) and the preserved bound fraction (pBF) are controls showing the depletion of RHOJ from the thioester-cleaving reagent and the near absence of non-specific binding of RHOJ to the resin (see Methods). Densitometric quantification of cBF/IF is shown (mean ± s.e.m.; $n = 3$ independent experiments; $^*P < 0.05$, one-sample *t*-test versus control). **v**, Left, autopalmitoylation enables endothelial GS to interact directly (or indirectly) with the Rho GTPase RHOJ and to sustain the palmitoylation, membrane localization and activity of RHOJ (reflected by GTP binding). RHOJ activity then sustains normal EC migration and lamellipodia formation, and keeps actin stress-fibre formation at levels that promote normal EC migration and vessel branching in vivo. Through mechanisms that are not completely understood, active RHOJ inhibits signalling of the RHOA/B/C–ROCK–(p)MLC pathway (itself known to promote stress-fibre formation). The relative contribution of a direct effect of RHOJ on migration versus the indirect effect through RHOA/B/C–ROCK–(p)MLC is yet to be determined. Reduced opacity of RHOA/B/C, ROCK and (p)MLC indicates reduced signalling of this pathway. Right, loss of endothelial GS renders RHOJ less active (visually reflected by fewer palmitoylated, membrane-bound RHOJ proteins), and reduces the inhibition of the RHOA/B/C–ROCK–(p)MLC pathway. The resulting excessive stress-fibre formation causes ECs to lose migratory capacity and reduces vessel branching in vivo. Dashed lines indicate reduced activity; the red cross indicates GS blockade; the question mark indicates unknown mechanisms. Exact *P* values are as follows: 0.0153 (**b**); 0.0334 (**e**); 2 versus 0 μM: 0.6327; 5 versus 0 μM: 0.2841; 10 versus 0 μM: 0.1090; 20 versus 0 μM: 0.0339; 40 versus 0 μM: 0.0034 (**h**); 1 versus 0.5 μg: 0.5806; 2 versus 0.5 μg: 0.0319; 4 versus 0.5 μg: 0.0037; 8 versus 0.5 μg: 0.0001; 16 versus 0.5 μg: 0.0001 (**i**); 0.0313 (**m**); RHOJ C3A versus RHOJ WT: 0.0001; RHOJ C11A versus RHOJ WT: 0.0001 (**q**); RHOJ C3A versus RHOJ WT: 0.0015; RHOJ C11A versus RHOJ WT: 0.0007 (**s**); 0.0051 (**t**); 0.0461 (**u**). Scale bar, 200 μm (**k**, **l**, **n**–**p**). For gel source images, see Supplementary Fig. 1.



Extended Data Fig. 8 | See next page for caption.

Extended Data Fig. 8 | Possible molecular model of GS autopalmitoylation. a, Structure of human GS, showing its bifunnel-shaped catalytic site. A schematic of the GS decamer is shown from the top and front views with individual subunits A and B labelled and coloured grey and green, respectively. On the right is a close-up of the bifunnel catalytic site that is formed between subunits A and B. The GS decamer has ten active sites, each located at the interface of two adjacent subunits. ATP enters from the top, whereas glutamate enters from below; manganese ions (Mn^{2+}) are shown as grey spheres. **b**, Molecular dynamics simulation of palmitoyl-CoA in the catalytic cleft of GS predicts that, whereas the head of palmitoyl-CoA is tightly bound to the adenine-binding site, the tail can point in opposing directions with respect to the principal axis of the protein. The most representative structures of the two alternative conformations (A and B) observed during the long molecular dynamics simulations for palmitoyl-CoA binding to GS (in blue, seen from two different perspectives) are shown in red (A, tail bending upwards) and green (B, tail bending downwards). **c**, Detailed view of conformation A, which is the main conformation. The sulfur atom of palmitoyl-CoA (which is immediately adjacent to the carbon on which the nucleophilic attack occurs) (coloured yellow) approaches the highly conserved cysteine 209 (also coloured yellow), with an interatomic distance (S–S) that, during the simulations, reversibly fluctuates between 3 and 8 Å. The hydrophobic tail positions itself along grooves characterized by the presence of hydrophobic residues. Colour coding is as follows: carbon, grey; nitrogen, blue; phosphorous, gold; oxygen, red. Cysteines and serines within 5 Å of the palmitoyl tail are highlighted in yellow and orange, respectively. The hydrophobic residues around the tail are shown in green. **d**, Detailed view of conformation B, in which the tail is found in a buried hydrophobic cleft, with the sulfur at a distance of 5 Å or less from the conserved serines 65 and 75 and the tail occupying the site of the GS inhibitor MSO. Details of the extensive steric clash between MSO and the secondary binding pose (B) observed in palmitoyl-CoA MD simulations are shown. Palmitoyl-CoA is represented as sticks, with standard atomic colours as stated in **c**. MSO is shown in cyan and its position is taken from entry 2QC8 of the Protein Data Bank. Cysteines and serines within 5 Å of the palmitoyl tail are highlighted in yellow and orange, respectively. The hydrophobic residues around the tail are shown in green. **e**, GS immunoblotting after streptavidin pull-down of biotin-azide-clicked lysates from 16C-YA (palmitoylation probe) labelled HEK-293T cells overexpressing wild-type *GLUL* or *GLUL* with a point mutation in C209. The input shows the level of GS overexpression. A representative blot from 4 independent experiments is shown. **f**, **g**, Quantification of total sprout length (**f**) and number of sprouts per spheroid (**g**) for control and $GLUL^{KD}$ ECs with or without overexpression of shRNA-resistant *GLUL* encoding the point mutation C209A (r$GLUL^{C209A\text{-}OE}$) (mean ± s.e.m.; $n = 4$ independent experiments; $^*P < 0.05$ versus control, one-way ANOVA with Dunnett's multiple comparison versus control). **h**, Schematic of protein autopalmitoylation. Upon binding of palmitoyl-CoA to the protein, free CoA (grey oval) is released and can be detected. **i**, Recombinant wild-type GS and GS with point mutations R324C and R341C were incubated with different concentrations of palmitoyl-CoA in a cell-free system at physiological pH. The amount of CoA released per minute was determined as a direct readout for protein autopalmitoylation. Data are mean ± s.e.m. of 3 (R324C and R341C) and 4 (WT) independent experiments. NS, $P > 0.05$; $^*P < 0.05$ according to two-way ANOVA comparing the entire dose–response to the dose–response of wild-type GS. **j**, Different amounts of recombinant wild-type, R324C and R341C GS were incubated with a fixed concentration of palmitoyl-CoA (40 μM), and the amount of CoA released per minute was determined as readout for autopalmitoylation. Data are mean ± s.e.m. of 4 (R324C and R341C) and 5 (wild-type) independent experiments. NS, $P > 0.05$; $^*P < 0.05$ according to two-way ANOVA comparing the entire dose–response to the dose–response of wild-type GS. The data for wild-type GS from **i** and **j** are also included in Extended Data Fig. 7 as stand-alone data, but are included here for comparison purposes. **k**, Boyden chamber migration for control, $GLUL^{KD}$, $GLUL^{KD}$ + r$GLUL^{OE}$, $GLUL^{KD}$ + r$GLUL^{R341C\text{-}OE}$ and $GLUL^{KD}$ + r$GLUL^{R324C\text{-}OE}$ ECs, all under mitomycin C-treatment (mean ± s.e.m.; $n = 3$ independent experiments; NS, $P > 0.05$; $^*P < 0.05$, one-way ANOVA with Dunnett's multiple comparison versus control). Exact *P* values: $GLUL^{KD}$ versus control: 0.0004; $GLUL^{KD}$ + r$GLUL^{C209A\text{-}OE}$ versus control: 0.0004 (**f**); $GLUL^{KD}$ versus control: 0.0001; $GLUL^{KD}$ + r$GLUL^{C209A\text{-}OE}$ versus control: 0.0001 (**g**); R324C versus WT: 0.8228; R341C versus WT: 0.7530 (**i**); R324C versus WT: 0.1331; R341C versus WT: 0.0003 (**j**); $GLUL^{KD}$ versus control: 0.0054; $GLUL^{KD}$ + r$GLUL^{OE}$ versus control: 0.8152; $GLUL^{KD}$ + r$GLUL^{R341C\text{-}OE}$ versus control: 0.3645; $GLUL^{KD}$ + r$GLUL^{R324C\text{-}OE}$ versus control: 0.2118 (**k**). For gel source images, see Supplementary Fig. 1.

Extended Data Table 1 | Weight, haematological and blood plasma parameters for six-week-old *Glul*^{vECKO} mice and control littermates

	Ctrl	*vECKO*
Hematological parameter		
WBCs ($x10^3$ μl^{-1})	4.42 ± 0.45	4.63 ± 0.49
neutrophils ($x10^3$ μl^{-1})	0.56 ± 0.13	0.40 ± 0.06
lymphocytes ($x10^3$ μl^{-1})	3.65 ± 0.41	4.05 ± 0.44
monocytes ($x10^3$ μl^{-1})	0.08 ± 0.02	0.03 ± 0.01*
eosinophils ($x10^3$ μl^{-1})	0.05 ± 0.01	0.06 ± 0.01
basophils ($x10^3$ μl^{-1})	0.08 ± 0.02	0.09 ± 0.02
platelets ($x10^3 \mu l^{-1}$)	766 ± 37	720 ± 72
RBCs ($x10^6$ μl^{-1})	6.89 ± 0.15	6.82 ± 0.14
hemoglobin (g dl^{-1})	10.63 ± 0.21	10.66 ± 0.23
hematocrit (%)	54.99 ± 1.17	55.00 ± 1.28
Liver parameter		
plasma AST (U l^{-1})	49.00 ± 4.00	49.50 ± 4.50
plasma ALT (U l^{-1})	26.50 ± 0.50	27.50 ± 4.50
plasma γ-GT (U l^{-1})	< 3	< 3
plasma bilirubin (mg dl^{-1})	< 0.18	< 0.18
Inflammation parameter		
CRP	< 0.3	< 0.3
Weight		
males (g)	18.80 ± 0.57	19.21 ± 1.22
females (g)	17.61 ± 0.80	17.77 ± 0.50

Values are mean ± s.e.m. of $n = 14$ (control; Ctrl) versus $n = 17$ (*Glul*^{vECKO}) mice. *$P = 0.0232$ versus control, Student's *t*-test. WBCs, white blood cells; RBCs, red blood cells; AST, aspartate amino transferase; ALT, alanine amino transferase; γ-GT, gamma-glutamyl transferase; CRP, C-reactive protein.

Extended Data Table 2 | Alignment of the amino acid sequence encompassing the C209 residue across different species

Species Name	Identity	Amino Acid Sequence (190 195 200 205 210 215 220 225 230)
Homo sapiens	100.00%	A G T N A E V M P - A Q W E F Q I G P **C** E G I S M G D H L W V A R F I L H R V C E
Canis lupus familiaris	98.39%	A G T N A E V M P - A Q W E F Q I G P **C** E G I S M G D H L W V A R F I L H R V C E
Macaca fascicularis	97.32%	A G T N A E V M P - A Q W E F Q I G P **C** E G I D M G D H L W V A R F I L H R V C E
Sus scrofa	96.51%	G G T N A E V M P - A Q W E F Q I G P **C** E G I D M G D H L W V A R F I L H R V C E
Bos taurus	96.25%	G G T N A E V M P - A Q W E F Q I G P **C** E G I D M G D H L W V A R F I L H R V C E
Acomys cahirinus	94.64%	T G T N A E V M P - A Q W E F Q I G P **C** E G I R M G D H L W I A R F I L H R V C E
Mus musculus	94.10%	T G T N A E V M P - A Q W E F Q I G P **C** E G I R M G D H L W V A R F I L H R V C E
Rattus norvegicus	92.49%	T G T N A E V M P - A Q W E F Q I G P **C** E G I R M G D H L W V A R F I L H R V C E
Cricetulus griseus	90.08%	T G T Y A E V K H - A Q W E F Q I G P **C** E G I R M G D H L W V A R F I L H R V C K
Gallus gallus	88.74%	G G T N A E V M P - A Q W E F Q V G P **C** E G I E M G D H L W I A R F I L H R V C E
Squalus acanthias	82.31%	S G T N A E V M A - A Q W E Y Q V G P **C** E G I Q M G D H L W I S R F I L H R V C E
Xenopus laevis	76.68%	C G T N A E V M P - S Q W E F Q V G P **C** E G I D M G D H L W M A R F I L H R V C E
Drosophila melanogaster	66.76%	S G E N A K V M P - A Q W E F Q V G P **C** E G I T M G D D L W M A R Y L L H R V A E
Panulirus argus	64.50%	S G T N A E V M P - A Q W E F Q V G P **C** E G I S I G D D L W M A R F L L H R I S E
Agaricus bisporus	58.18%	A G T N A E V M P - A Q W E F Q I G P **A** - G I K A C D D L W V S R Y I L Q R I A E
Hebeloma cylindrosporum	57.70%	S G I N A E V M P - S Q W E Y Q V G P **C** H G I E M G D H L W I S R F L L H R V A E
Suillus bovinus	57.69%	S G V N G E V L P - S Q W E Y Q V G P **V** E G I A M G D Q M W M S R Y L M Y R V A E
Cryptococcus neoformans var. neoformans B-3501A	57.46%	S G T N A E V M P - A Q W E F Q V G P **C** D G I E M G D H L W L A R F L L H R V S E
Aspergillus nidulans FGSC A4	57.34%	S G I N A E V M P - S Q W E F Q V G P **C** E G I S M G D H L W M A R Y L L V R I A E
Amanita muscaria	57.06%	S G I N A E V M P - S Q W E F Q V G P **C** E G I S M G D H L W M A R Y L L I R V A E
Neurospora crassa OR74A	56.98%	S G I N A E V M P - S Q W E F Q V G P **C** T G I E M G D H L W M A R F L L L R I G E
Tuber borchii	56.98%	S G I N A E V L P - S Q W E F Q V G P **C** V G I E M G D H L W I S R Y L L H R V A E
Caenorhabditis elegans	56.50%	S G I N A E V M P - S Q W E F Q V G P **C** E G I S M G D H L W M A R Y L L V R I A E
Candida glabrata CBS 138	56.39%	S G T N A E V M P - A Q W E F Q V G P **C** V G I E M G D H L W L A R F L L A R I A E
Colletotrichum gloeosporioides	55.24%	S G I N A E V M P - G Q W E F Q V G P **S** V G I S A G D E L W V A R Y I L E R I T E
Kluyveromyces lactis NRRL Y-1140	55.24%	S G I N A E V M P - G Q W E F Q V G P **S** I G I S A G D E L W V A R Y I L E R I T E
Debaryomyces hansenii CBS767	55.06%	S G I N G E V M P - G Q W E F Q V G P **S** V G I S A G D E I W A A R Y I L E R I T E
Yarrowia lipolytica CLIB122	55.04%	F G T N A E V T P - G Q W E F Q I G T **C** E G I D M G D Q L W M S R Y I L H R V A E
Eremothecium gossypii ATCC 10895	54.96%	G G A N G E V M P - G Q W E F Q I S P **T** V G I G A G D Q L W V A R Y I L E R I T E
Saccharomyces cerevisiae S288C	54.78%	S G T N A E V M P - A Q W E Y Q V G P **C** L G I E M G D Q L W V S R F F L A R I T E
Fusarium phaseoli	54.52%	S G I N G E V M P - G Q W E F Q V G P **V** E G I S S G D Q V W V A R Y L L E R I T E
Medicago sativa	54.49%	S G I N G E V M P - G Q W E F Q V G P **A** V G I S A G D E L W V A R Y I L E R I T E
Schizosaccharomyces pombe 972h-	54.39%	S G T N A E V M P - A Q W E Y Q V G P **C** T G I A M G D Q L W I S R F F L H R V A E
Fusarium fujikuroi	54.21%	S G I N G E V M P - G Q W E F Q V G P **S** V G I S A G D Q V W V A R Y I L E R I T E
Oryza sativa Japonica Group	54.21%	S G I N G E V M P - G Q W E F Q V G P **A** V G I S A G D E L W V A R Y I L E R I T E
Lupinus luteus	54.08%	S G I N G E V M P - G Q W E F Q V G P **S** V G I S A G D E I W V A R Y I L E R I T E
Pinus sylvestris	53.93%	S G I N G E V M P - G Q W E F Q V G P **S** V G I S A G D E I W V A R Y I L E R I T E
Arabidopsis thaliana	53.76%	S G I N A E V M P - S Q W E F Q V G P **C** E G I S M G D Q L W M A R Y F L H R V A E
Pisum sativum	53.76%	S G I N A E V M P - S Q W E Y Q V G P **C** A G I E M G D Q L W M S R F L L H R I A E
Phaseolus vulgaris	53.65%	S G I N G E V M P - G Q W E F Q V G P **S** V G I S S G D Q V W V A R Y I L E R I T E
Lactuca sativa	53.63%	S G I N G E V M P - G Q W E F Q V G P **S** V G I A A A D Q I W V A R Y I L E R I T E
Vigna aconitifolia	53.50%	S G I N G E V M P - G Q W E F Q V G P **S** V G I S A A D E L W C A R F I M E R I T E
Zea mays	53.37%	S G I N G E V M P - G Q W E F Q V G P **S** V G I S A G D E V W A A R Y I L E R I T E
Lotus japonicus	53.37%	S G I N G E V M P - G Q W E F Q V G P **S** V G I S A A D E L W V A R Y I L E R I T E
Nicotiana plumbaginifolia	53.24%	S G I N G E V M P - G Q W E F Q V G P **S** V G I S A G D E I W A A R Y I L E R I T E
Hordeum vulgare	52.94%	S G I N G E V M P - G Q W E F Q V G P **S** V G I S A G D E I W A A R Y I L E R I T E
Vitis vinifera	52.81%	S G I N G E V M P - G Q W E F Q V G P **S** V G I S A G D E V W A A R Y I L E R I T E
Daucus carota	52.25%	S G T N G E V M P - G Q W E F Q V G P **T** V G I A A A D Q V W V A R Y I L E R I T E
Glycine max	52.25%	S G I N G E V M P - G Q W E F Q V G P **T** V G I S A G D Q V W V A R Y I L E R I T E
Alnus glutinosa	50.94%	S G T N G E V M P - G Q W E F Q V G P **S** V G I D A G D H V W C A R Y L L E R I T E
Brassica napus	50.40%	S G T N G E V M P - G Q W E F Q V G P **S** V G I E A G D H V W C A R Y L L E R I T E
Chlamydomonas reinhardtii	50.13%	S G T N G E V M P - G Q W E Y Q V G P **S** V G I E A G D H I W I S R Y I L E R I T E
Lupinus angustifolius	49.60%	S G T N G E V M P - G Q W E Y Q V G P **S** V G I E A G D H I W A S R Y I L E R I T E
Streptomyces viridochromogenes	49.33%	S G V N A E V L P - S Q W E Y Q V G P **C** E G I T M G D H M W M S R Y I M Y R V C E
Acanthamoeba polyphaga mimivirus	48.53%	S G I N A E V M P - G Q W E F Q V G A **L** P P L E V S D H M W V A R W L L H R V A E
Streptomyces hygroscopicus	48.01%	E G T N A E V M M - G Q W E F Q I G V **L** P A P A I G D Q I W L G R W L L H R I A E
Frankia alni	46.92%	S G I N A E V M P - G Q W E Y Q I G P **V** G P L A L G D E V M L S R W L L H R L G E
Dunaliella salina	45.04%	S G T N L E V A P - G Q H E F Q I G P **V** E G I D A A D Q L W I A R F I L E K I S E
Bradyrhizobium diazoefficiens USDA 110	43.56%	E G I N A E V A K - G Q W E F Q I F G **K** G S K K A A D Q I W M A R Y L L Q R L T E
Rhizobium leguminosarum bv. phaseoli	42.55%	E G I N A E V A K - G Q W E F Q I F G **K** G S K K A A D E V C V A R Y L L V R L T E
Sinorhizobium meliloti	42.44%	E G I N A E V A K - G Q W E F Q I F G **K** G S K K A A D E M W M A R Y L M L R L T E
Helicobacter pylori 26695	25.21%	E K H H H E V A T A G Q N E I A C Q F **N** T L V K K A D E I Q V Y K Y V V H N V A H
Salmonella enterica subsp. enterica serovar Typhimurium str. LT2	23.32%	E A H H H E V A T A G Q N E V A T R F **N** T M T K K A D E I Q I Y K Y V V H N V A H
Methanocaldococcus jannaschii DSM 2661	23.06%	E A H H H E V A T A G Q N E V A T R F **N** T M T K K A D E I Q I Y K Y V V H N V A H
Azotobacter vinelandii	22.52%	E A H H H E V A T A G Q N E V A T R F **N** T M T K K A D E T Q I Y K Y V V Q N V A H
Azorhizobium caulinodans ORS 571	22.25%	F V V H H E V A Q A - Q G E V G V K F **G** D L V E A A D N V Q K L K Y V V K M V A H
Thermotoga maritima MSB8	21.98%	E K H H H E V A - A S Q H E L G I K F **D** T L V R T G D N M Q Y Y K Y V V H N V A H
Azospirillum brasilense	21.98%	E V H H H E V A T A G Q C E I G V R C **N** T L V K K A D E V L L L K Y A V Q N V A H
Pseudomonas aeruginosa PAO1	21.72%	E K H H H E V A - S A Q H E L G L K F **G** Q L V T M A D H L Q V Y K Y C I H Q V A N
Proteus hauseri	21.72%	E A H H H E V A T A G Q N E I A T K F **N** T L T L K A D E T Q I Y K H V V Q N V A L
Methylococcus capsulatus str. Bath	21.72%	E V L H H E V G K - A Q H E I D F R Y **D** E A L K T A D N I V S F K Y I T K A V A E
Microchaete diplosiphon	21.72%	E V L H H E V G K - A Q H E I D F R Y **D** E A L K T A D N I V S F K Y I T K A V A E
Vibrio cholerae O1 biovar El Tor str. N16961	21.45%	E A H H H E V A T A G Q N E I A T R F **N** T L T T K A D E I Q I Y K Y V V H N V A H
Methanococcus voltae	21.45%	E K H H H E V A - S A Q H E L G M K F **D** T L T L M A D H L Q I Y K Y C I H Q V A H
Neisseria gonorrhoeae	21.45%	E A S H H E V A R - G Q H E I N F K Y **D** D A L T T A D N I A T F R A V V R A V A E
Bacillus cereus	21.45%	E V L H H E V G K - A Q H E I D F R Y **D** E A L K T A D N I I S F K Y I V K A V A E
Synechocystis sp. PCC 6803 substr. Kazusa	21.18%	E V H H H E V A T A G Q N E I G V K F **N** T L V A K A D E V Q T L K Y C V H N V A D
Vibrio alginolyticus	21.18%	E V H H S E V G T G S Q M E I G T R F **A** T L V K R A D Q T Q D M K Y V I Q N V A H
Pseudomonas taetrolens	21.18%	E K H H H E V A S G G Q N E L G I K F **D** K L V N S A D N L M I Y K Y V I K N V A K
Sulfolobus acidocaldarius DSM 639	21.18%	E V L H H E V G K - A Q H E I D F R H **D** E A L K T A D N I V S F K Y I V K A V A E
Staphylococcus epidermidis ATCC 12228	21.18%	E V H H H E V A G Q G Q N E I G T R F **S** T L V Q R A D W T Q L Q K Y V I Q N V A H
Rhodobacter capsulatus	21.18%	E A H H H E V A T A G Q N E I A T R F **N** T L T T K A D E T Q I Y K Y V V Q N V A Y
Bacillus subtilis subsp. subtilis str. 168	20.91%	E V H H H E V A T A G Q N E I G V K F **N** T L V K K A D E V Q T L K Y V V H N V A D
Staphylococcus aureus subsp. aureus MRSA252	20.91%	E V H H H E V A T A G Q H E I G V G F **N** T L T P R R M R C K I L K Y V V H N V A -
Halobacterium salinarum NRC-1	20.91%	E A S H H E V A E - G Q H E I N F T Y **D** D A L S T A D N V A T F R S V V R A I A A
Haemophilus influenzae Rd KW20	20.64%	E K H H H E V A T G G Q C E L G F R F **G** K L I E A A D W L M I Y K Y V I K N V A K
Nostoc sp. PCC 7120	20.38%	E V H H H E V A T A G Q N E I G V S F **N** T L V A K A D E V Q T L K Y C V H N V A D
Pyrococcus horikoshii OT3	20.38%	E K Q H H E V A T G G Q C E L G F R F **G** K L I E A A D W L M T Y K Y V I K N V A K
Sinorhizobium meliloti 1021	20.38%	E R Q H H E V G T A G Q A E I N Y K F **N** T L L A A A D D L M L F K Y I V K N V A W
Ralstonia eutropha JMP134	20.11%	E A S H H E V A P - G Q H E V D F K F **D** N A L K T A D S V I T F K M T I K N I A K
Pyrococcus woesei	20.11%	D K H H H E V A T C - Q H E L G L I F **G** G L T E Q A D N I L K Y K Y V I H N V A H
Pyrococcus abyssi GE5	20.11%	E K H H H E V A T G G Q N E L G F R F **A** T L L K A A D Y L M T Y K Y V I K N V A R
Methanothermobacter thermautotrophicus str. Delta H	20.11%	E V L H H E V G K - A Q H E I D F R Y **D** E A L R T A D N I V S F K H V V K A V A E
Pasteurella multocida subsp. multocida str. Pm70	19.84%	E A S H H E V A P - G Q H E I D F K Y **A** G A V R S C D D I Q T F K L V V K T I A R
Aquifex aeolicus VF5	19.84%	E K H H H E V A - A A Q H E L G V K F **D** A L V R N A D K M Q I Y K Y V V H Q V A N
Clostridium saccharobutylicum	19.84%	E A S H H E V A E - G Q N E I D F K Y **G** D A L T T A D N I M T F K L V V K S I A Q
Synechococcus sp. PCC 7002	19.57%	E A A H H E V A T A G Q G E I D F R F **S** T L A D T A D K V Q V L K Y V T K N I A S
Pyrococcus furiosus DSM 3638	19.57%	E A S H H E V A P - G Q H E I D F K Y **A** D A V T A C D N I Q T F K L V V K T I A R
Rhodobacter sphaeroides	19.57%	D K H H H E V A T C - Q H E L G L I F **G** G L T E Q A D N I L K Y K Y V I H N V A G
Mycobacterium smegmatis str. MC2 155	19.57%	E A S H H E V A E - G Q H E V D F K F **A** D A V K T A D S V V T F K T T I K T L A A
Escherichia coli	28.00%	E A H H H E V A T A G Q N E V A T R F **N** **T** M T K K A D E I Q I Y K Y V V H N V A H

Multiple sequence alignment showing the conservation of amino acid C209 (in red) in GS across different species. Here the sequence alignment of 41 residues surrounding this cysteine is shown for up to 100 of the closest sequence identity matches with the GS of *H. sapiens* obtained with BLAST from the UniProtKB/Swiss-Prot database. C209 is mostly conserved across species, and when not conserved it is often substituted by residues (Ser or Thr) that can (in theory) be palmitoylated as well. In *E. coli* (shown at the bottom), for example, a Thr is found at the structurally equivalent position (T210, highlighted in yellow). If several GS isoforms are known for the same species, only the one with the highest percentage identity to that of *H. sapiens* is shown.

ARTICLE

https://doi.org/10.1038/s41586-018-0469-4

Malaria parasite translocon structure and mechanism of effector export

Chi-Min Ho[1,2,3], Josh R. Beck[4,5,6], Mason Lai[2], Yanxiang Cui[7], Daniel E. Goldberg[4,5], Pascal F. Egea[1,3]* & Z. Hong Zhou[1,2,7]*

The putative *Plasmodium* translocon of exported proteins (PTEX) is essential for transport of malarial effector proteins across a parasite-encasing vacuolar membrane into host erythrocytes, but the mechanism of this process remains unknown. Here we show that PTEX is a bona fide translocon by determining structures of the PTEX core complex at near-atomic resolution using cryo-electron microscopy. We isolated the endogenous PTEX core complex containing EXP2, PTEX150 and HSP101 from *Plasmodium falciparum* in the 'engaged' and 'resetting' states of endogenous cargo translocation using epitope tags inserted using the CRISPR–Cas9 system. In the structures, EXP2 and PTEX150 interdigitate to form a static, funnel-shaped pseudo-seven-fold-symmetric protein-conducting channel spanning the vacuolar membrane. The spiral-shaped AAA+ HSP101 hexamer is tethered above this funnel, and undergoes pronounced compaction that allows three of six tyrosine-bearing pore loops lining the HSP101 channel to dissociate from the cargo, resetting the translocon for the next threading cycle. Our work reveals the mechanism of *P. falciparum* effector export, and will inform structure-based design of drugs targeting this unique translocon.

Malaria has devastated major civilizations since the dawn of humanity, and remains a considerable burden to society; it is responsible for more than 200 million cases and nearly half a million deaths each year[1]. This infectious disease is caused by *Plasmodium* parasites, which invade and reproduce within human erythrocytes, inducing the clinical symptoms of malaria[2,3]. These parasites export hundreds of effector proteins that extensively remodel host erythrocytes, which have limited capacity for biosynthesis[4–6]. Collectively known as the exportome, these proteins create the infrastructure necessary to import nutrients, export waste, and evade splenic clearance of infected erythrocytes[7]. Most of these proteins bear a five-residue motif called the *Plasmodium* export element (PEXEL)[8–10]. The malaria parasite conceals itself inside a parasitophorous vacuole, which is derived from invagination of the host cell plasma membrane during invasion[11] (Fig. 1a). Following secretion into the parasitophorous vacuole, proteins destined for export are unfolded and transported across the parasitophorous vacuole membrane (PVM) into the host cell in an ATP-dependent process[12,13]. To accomplish this, it has been hypothesized that the parasite has evolved a unique membrane protein complex, the *Plasmodium* translocon of exported proteins (PTEX)[14]. PTEX is the only known point of entry to the host cell for exported proteins and is therefore an attractive drug target, as disrupting PTEX blocks delivery of key virulence determinants and induces parasite death[15,16].

PTEX has been proposed to be a membrane protein complex larger than 1.2 MDa, with a core composed of the HSP101 ATPase and two novel proteins, PTEX150 and EXP2[14,17] (Fig. 1a). HSP101 belongs to the class 1 Clp/HSP100 family of AAA+ ATPases, PTEX150 has no known homologues beyond the *Plasmodium* genus, and EXP2 is a PVM protein[14,18] that is conserved among vacuole-dwelling apicomplexans[19]. All three core components are essential for protein export and parasite survival[15,16,20]. A model of PTEX-mediated translocation has been proposed, in which HSP101 unfolds and threads proteins through an oligomeric EXP2 transmembrane channel spanning the PVM, with PTEX150 having a structural role between EXP2 and HSP101[14–17]. However, without structural information, the global architecture of PTEX, the stoichiometry of its components and its molecular mechanism have remained unclear.

In this study, we purify PTEX directly from the human malaria parasite *P. falciparum* and determine the structure of the complex in multiple functional states at near-atomic resolution using cryo-electron microscopy (cryo-EM). Our atomic models reveal the architecture and mechanism of this unique translocon and open a path towards development of novel therapeutics against this promising anti-malarial drug target.

Architecture of the PTEX core complex

To purify PTEX from *P. falciparum*, we used CRISPR–Cas9 editing to introduce a 3×Flag epitope tag on the C terminus of endogenously expressed HSP101 (Extended Data Fig. 1a–c) and purified the endogenously assembled PTEX core complex directly from *P. falciparum* cultured in human erythrocytes (Extended Data Fig. 1d–f). Cryo-EM analysis yielded two distinct conformations of PTEX particles, one extended (195 Å) and the other compact (175 Å) (Fig. 1b, c, Extended Data Table 1). Endogenous cargo polypeptide densities are visible in the central pore of HSP101 in both structures (Fig. 1b, c, Extended Data Figs. 2, 3). On the basis of differences in the arrangement of HSP101 subunits relative to the cargo between the two conformations, we designated them as the 'engaged' and 'resetting' states, respectively. Both maps are at near-atomic resolution, varying from 3.0–3.6 Å in the transmembrane and core regions to 5–8 Å in the periphery (Fig. 1b, c, Extended Data Fig. 4). As most regions exhibit clear sidechain densities throughout both maps (Fig. 1b, c, Extended Data Fig. 5, Supplementary Videos 1, 2) we were able to build de novo atomic models of the three constituent proteins for both conformational states (Fig. 1d, e). Each model contains 20 subunits with 6,898 modelled amino acid residues. Models of all subunits were built independently, as their conformations varied.

[1]The Molecular Biology Institute, University of California, Los Angeles, CA, USA. [2]Department of Microbiology, Immunology & Molecular Genetics, University of California, Los Angeles, CA, USA. [3]Department of Biological Chemistry, David Geffen School of Medicine, University of California, Los Angeles, CA, USA. [4]Department of Medicine, Division of Infectious Diseases, Washington University School of Medicine, St. Louis, MO, USA. [5]Department of Molecular Microbiology, Washington University School of Medicine, St. Louis, MO, USA. [6]Department of Biomedical Sciences, Iowa State University, Ames, IA, USA. [7]California NanoSystems Institute, University of California, Los Angeles, CA, USA. *e-mail: pegea@mednet.ucla.edu; hong.zhou@UCLA.edu

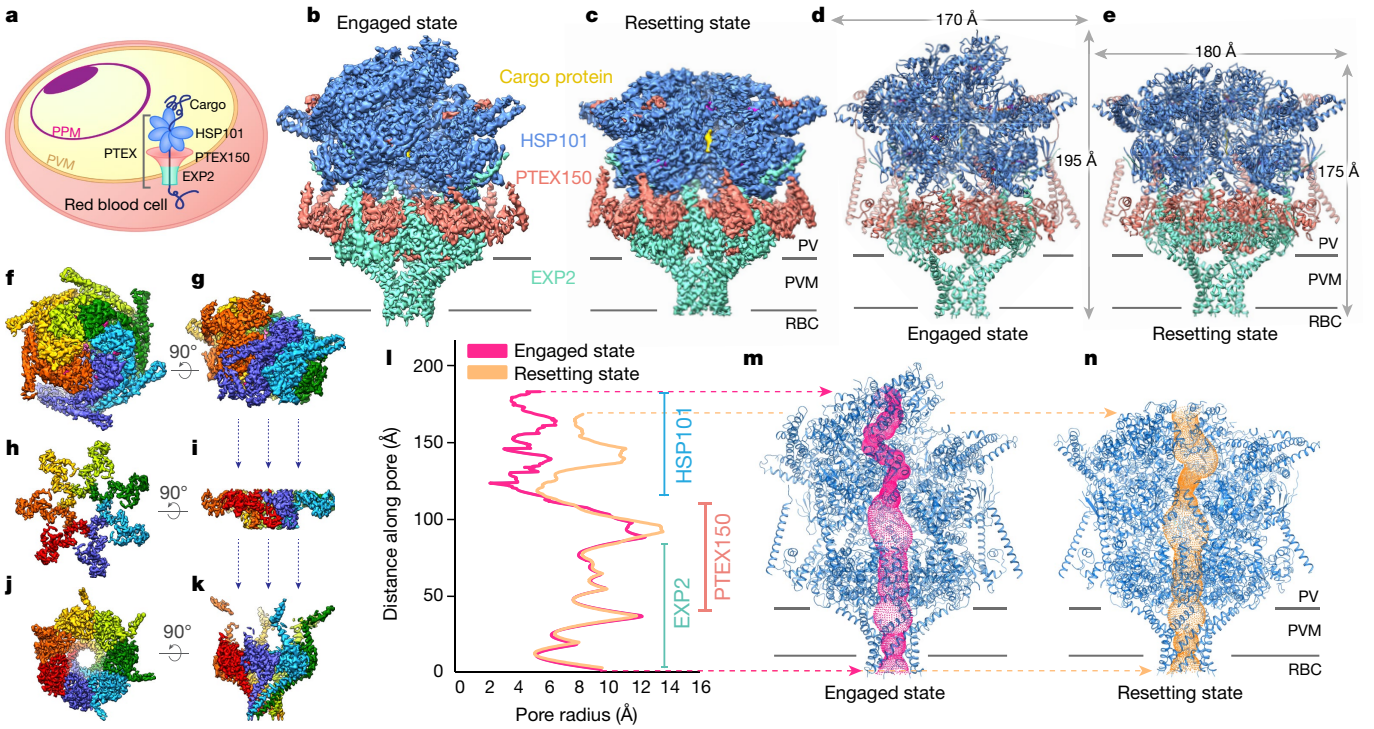


Fig. 1 | Global architecture of the PTEX core complex in two cargo-bound states. a, Schematic of a human erythrocyte infected with *P. falciparum*. PPM, parasite plasma membrane; PVM, parasitophorous vacuole membrane. **b**–**e**, Cryo-EM maps (**b**, **c**) and atomic models (**d**, **e**) of the PTEX core complex. Horizontal lines represent the PVM bilayer, estimated on the basis of the detergent belt density, which is visible at lower thresholds (see Extended Data Fig. 7). PV, parasitophorous vacuole; RBC, red blood cell. **f**–**k**, Top and side views of the HSP101 (**f**, **g**), PTEX150 (**h**, **i**) and EXP2 (**j**, **k**) cryo-EM maps, coloured by protomer. **l**–**n**, Pore radius (**l**) and protein-conducting channel (**m**, **n**) calculated using HOLE[40].

Both structures show that PTEX is a tripartite membrane protein complex with a 6:7:7 stoichiometry and a calculated mass of 1.6 MDa, composed of a hexameric HSP101 protein-unfolding motor tethered to a PVM-spanning, pseudosymmetric funnel formed by seven protomers of EXP2 interdigitating with seven protomers of PTEX150 (Fig. 1d–k, Supplementary Video 3). Two transiently associated[21] accessory proteins, PTEX88 and TRX2[14], are not seen in our structures. At the PVM, each EXP2 monomer contributes a single transmembrane helix to form a seven-fold (C7)-symmetric protein-conducting channel spanning the membrane (Fig. 1j, k). Six HSP101 protomers are tethered on top of the PTEX150–EXP2 funnel in a hexameric right-handed spiral, with a gap between the bottom-most and top-most protomers (Fig. 1f, g, Supplementary Video 4). The HSP101 hexamer is oriented such that a single unbroken channel extends from the top of the HSP101 hexamer to the bottom of the heptameric EXP2 transmembrane pore (Fig. 1l–n, Extended Data Fig. 2d). The most constricted point along the channel occurs in HSP101, measuring 4 Å and 10 Å in diameter in the engaged and resetting states, respectively (Fig. 1l). The seventh EXP2 and PTEX150 protomers are situated under the gap between HSP101 protomers 1 and 6, accommodating the remarkable symmetry mismatch between the asymmetric HSP101 hexamer and the pseudo-seven-fold-symmetric PTEX150–EXP2 tetradecamer (Fig. 1f–k, Extended Data Fig. 2e–j). Analyses of our PTEX150 and EXP2 structures with four commonly used structural similarity search programs[22–25] found no consistent structural similarities to any known proteins, including the pore-forming toxin haemolysin E, with which EXP2 was previously speculated to share structural homology[14]. Below, we describe the structural details of the individual proteins in the engaged state, followed by a comparison of the two states that suggests a mechanism of translocation.

EXP2 forms a protein-conducting channel across PVM

Residues G27–S234 of EXP2 are well-resolved in our structure, accounting for 80% of the mature protein (Extended Data Fig. 6a). EXP2 is a single-pass transmembrane protein consisting of a kinked 60 Å N-terminal transmembrane helix followed by a globular body domain and ending in an assembly domain composed of a linker helix followed by the assembly strand (Fig. 2a, b). The body domain contains five helices, B1—B5, which are stabilized by an intraprotomer C113—C140 disulphide bond (Fig. 2c).

Seven EXP2 protomers (which we labelled A—G) oligomerize to form a funnel-shaped C7-pseudosymmetric 216-kDa heptamer spanning the PVM (Fig. 2d, e). The transmembrane domain and body helices B1–B3 are symmetric throughout all seven protomers (Extended Data Fig. 3a, b). This symmetry is broken by inter-protomer conformational variations in body helices B4–B5 and the assembly domain, which stretch upwards in some protomers to maintain contacts with the asymmetric HSP101 hexamer located above the EXP2 funnel. This variation is most pronounced in EXP2 protomers F and G (Extended Data Fig. 3a, b).

In the EXP2 heptamer, the amphipathic transmembrane helices twist slightly around each other, creating a 37 Å-long C7-symmetric protein-conducting channel that spans the PVM and forms the stem of the funnel (Fig. 2d, e). The membrane-facing surface of the EXP2 channel is coated with hydrophobic residues, whereas the inner surface is lined with charged and polar residues, creating an aqueous pore (Fig. 2e). The body domains, positioned in a wider ring on top of the transmembrane channel on the vacuolar face of the PVM, form the mouth of the funnel. This orientation is consistent with previous analyses of EXP2 topology[14,20]. Furthermore, a detergent belt is clearly visible in 2D class averages and density maps (Extended Data Figs. 7, 8), defining the residues in the transmembrane domain that would be buried in the PVM. A ring of positively charged residues where the stem meets the mouth of the funnel is positioned where it can interact with the negatively charged phosphates of the membrane surface (Extended Data Fig. 8a).

PTEX150 forms an adaptor between HSP101 and EXP2

Of the 993 residues in PTEX150, S668—D823 are well-resolved in the structure, and form a hook with a shaft (Fig. 3a, b). The hook

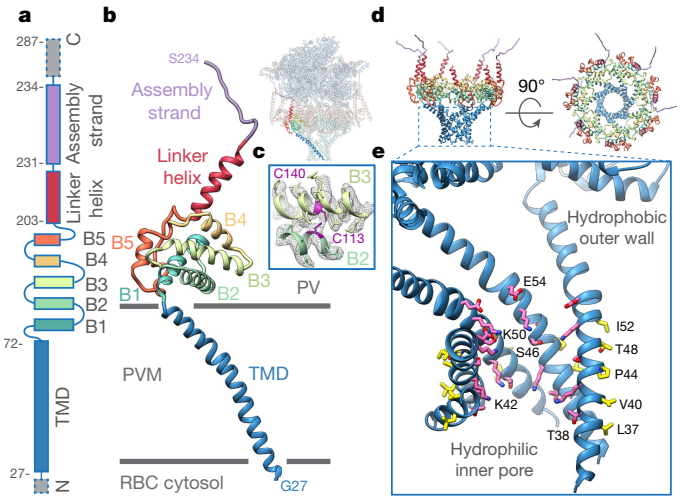


Fig. 2 | EXP2 forms a heptameric pseudo-symmetric PVM-spanning pore. a, **b**, Linear schematic (**a**) and ribbon diagram (**b**) of the EXP2 monomer in the engaged state. Dashed grey boxes represent unmodelled regions. Inset, one EXP2 monomer (coloured) within the PTEX complex. TMD, transmembrane domain. **c**, Density (mesh) and model of the C113–C140 disulphide bond. **d**, EXP2 heptamer, coloured as in **b**. **e**, Cutaway of the EXP2 transmembrane channel with hydrophilic residues (pink) lining the inner protein-conducting pore and hydrophobic residues (yellow) on the outer, membrane-facing surface.

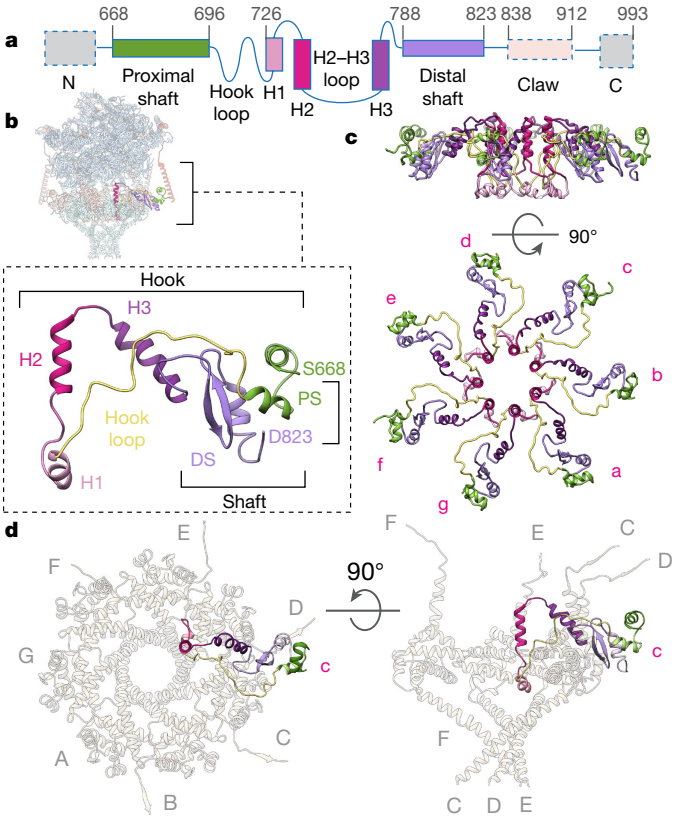


Fig. 3 | PTEX150 forms a heptameric flange-shaped adaptor between EXP2 and HSP101. a, **b**, Linear schematic (**a**) and ribbon diagram (**b**) of the PTEX150(668–823) monomer in the engaged state. Dashed grey boxes represent unmodelled regions. Inset, one PTEX150(668–823) monomer (coloured) within the PTEX complex. PS, proximal shaft; DS, distal shaft. **c**, The PTEX150(668–823) heptamer, coloured as in **b**. **d**, Views showing how one PTEX150(668–823) monomer hooks into the top of the EXP2 funnel.

domain consists of three short helices (H1–H3), which are connected by several long loops. Directly N-terminal and C-terminal to the hook domain, the shaft is composed of proximal and distal shaft domains (Fig. 3a, b). The remaining 80% of PTEX150—not visible in our structures—is predicted to be intrinsically disordered (in this region the average disorder tendency score in IUPred[26,27] is 0.83, with scores above 0.5 indicating disorder), unlike the rigid structured core of PTEX150(S668–D823) (with an average disorder tendency score of 0.42, indicating ordered structure), suggesting that this 80% of the protein is too mobile to be observed and may be flexibly arranged outside the stable PTEX core.

Seven PTEX150(S668–D823) hooks (which we labelled a–g) oligomerize, forming a flange-shaped C7-pseudosymmetric heptamer (Fig. 3c) that fits into the mouth of the EXP2 channel. Each hook lies in the groove between adjacent EXP2 body domains, and the tip of the hook curls down into the mouth of the EXP2 pore (Fig. 3d). A vertical, heptameric ring of H2 helices sits in the mouth of the EXP2 funnel, forming a conduit between the hexameric HSP101 and heptameric EXP2 central pores (Extended Data Fig. 2g–j). In this way, PTEX150(S668–D823) serves as an adaptor between HSP101 and EXP2, providing a continuous protected path for unfolded cargo.

Endogenous cargo is observed in the channel of HSP101

Class 1 Clp–HSP100 AAA+ ATPases are highly conserved hexameric protein unfoldases that are associated with diverse functions; they are known to thread polymeric substrates through a central pore[28,29]. HSP101 is a 598-kDa hexamer, exemplifying the canonical class 1 Clp–HSP100 domain architecture[30,31], with a substrate-binding N-terminal domain[32] followed by two AAA+ nucleotide-binding domains (NBD1 and NBD2), each containing a cargo-binding pore loop (L1 and L2, respectively) that extends into the central pore (Fig. 4a, b). Additionally, HSP101 contains a C-terminal domain and a coiled-coil middle domain insertion in the C-terminal end of NBD1 (Fig. 4a, b).

Unlike class 2 HSP100s [33], class 1 HSP100s form three-tiered hexamers, in which the N-terminal domains, NBD1s and NBD2s form the top, middle and bottom tiers, respectively[30,31]. In our engaged state structure, the NBD1 and NBD2 tiers are arranged in a right-handed ascending spiral[30,31,34] (Fig. 4c). A layer of weaker density above the NBD1 tier may correspond to the N-terminal domains, which are likely to be dynamic (Extended Data Fig. 8b). The middle domains encircle the upper NBD1 tier. The central pore of the spiral is lined with pore loops bearing tyrosines in a spiral staircase pattern. The tyrosine sidechain densities intercalate with a 45 Å-long density that is clearly visible in the middle of the chaperone pore (Fig. 4c, d, Supplementary Video 5), which closely resembles unfolded cargo polypeptide densities reported in recently published cryo-EM structures of homologous HSP100s bound to cargo[31,33] (Extended Data Fig. 2a, d). The unfolded PTEX cargo polypeptide chain modelled into this 45 Å density matches very closely (root mean square deviation (r.m.s.d.) of 1.09–1.25 Å) with the unfolded cargo polypeptides in these cargo-bound homologue structures (Extended Data Fig. 2a–c).

Key interactions for PTEX assembly and function

The three PTEX components share extensive binding interfaces; here we describe the two most noteworthy interactions. In each EXP2 protomer A–F, the assembly strand augments the C-terminal domain β-sheet in the HSP101 protomer situated directly above it (Fig. 5a, b). Protomer G occupies the space below the gap between HSP101 protomers 1 and 6 (Fig. 1f–k). This hydrogen bond interaction tethers the HSP101 hexamer to the transmembrane funnel, positioning the central pore exit directly above the entrance to the PTEX150–EXP2 pore. We hypothesized that this interaction is essential for assembly of the PTEX core complex, and that the complex must be stably assembled to be active. We tested this using genetic functional complementation in live *P. falciparum*.

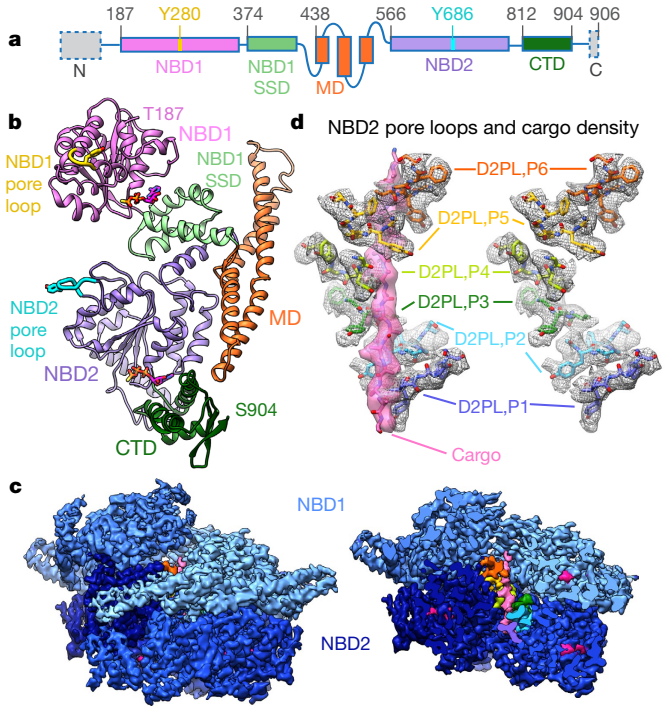


Fig. 4 | Endogenous cargo bound in the channel of the HSP101 hexamer. a, **b**, Linear schematic (**a**) and ribbon diagram (**b**) of the HSP101 monomer in the engaged state. MD, middle domain; CTD, C-terminal domain. SSD, small subdomain. **c**, Side view of the full (left) and bisected (right) cryo-EM map of the HSP101 hexamer. NBD1 and NBD2 rings are coloured with light (NBD1) and dark (NBD2) blue gradients to emphasize the right-handed spiral shape of the hexamer. The bisected map shows NBD2 pore loop densities coloured by protomer, ATPγS (magenta) and the cargo density (light pink). **d**, Enlarged side view of the atomic models of the HSP101 NBD2 pore loops and unfolded cargo polypeptide backbone, shown with densities. NBD2 pore loops are coloured as in **c** and labelled by protomer (for example, D2PL,P1 is NBD2 pore loop, protomer 1). Vertical distances between pore loop tyrosines in D2PL,P1–D2PL,P6 are 6.52 Å, 6.28 Å, 6.38 Å, 6.96 Å and 6.12 Å, respectively.

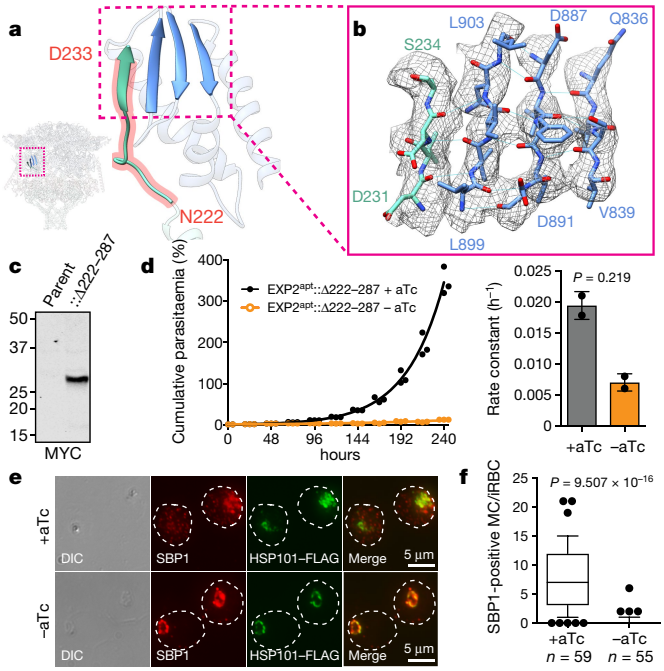


Fig. 5 | Interactions essential for PTEX function. a, **b**, Ribbon (**a**) and stick (**b**) models of the HSP101 C-terminal domain β-sheet augmented by the EXP2 assembly strand, shown with corresponding cryo-EM density (mesh). The segment outlined in red was truncated in functional complementation assays. **c**, Western blot of lysates from *P. falciparum* expressing EXP2 under aptamer control (EXP2^{apt}) complemented with EXP2(Δ222–287)–3×Myc (predicted molecular weight 27.8 kDa after signal peptide cleavage). For blot source data, see Supplementary Fig. 1. **d**, Growth analysis of EXP2^{apt}::EXP2(Δ222–287)–3×Myc. *P. falciparum* were grown with or without anhydrotetracycline (aTc) to maintain or knockdown endogenous EXP2 expression, respectively. One experiment performed with three technical replicates is shown. Bars show mean exponential growth rate constant (h^{-1}) determined from the fit of the two independent experiments and error bars indicate s.d. **e**, Immunofluorescence assay detecting exported protein SBP1 and HSP101–3×FLAG (a marker of the parasitophorous vacuole) in *P. falciparum* expressing EXP2^{apt}::EXP2(Δ222–287)–3×Myc that were allowed to develop with or without aTc for 24 h after invasion. The dashed line indicates the traced boundary of the red blood cell. DIC, differential interference contrast image. **f**, Quantification of SBP1 export immunofluorescence assays. Data are pooled from two independent experiments; *n* is the number of individual parasite-infected RBCs. Boxes and whiskers delineate 25th to 75th and 10th to 90th percentiles, respectively. All *P* values are determined by an unpaired, two-sided Student's *t*-test. All data shown represent two independent experiments.

Knockdown of EXP2 produces a lethal defect in *P. falciparum* growth and export that can be rescued by a mutant version of EXP2 that lacks the C-terminal 54 residues[20]. Therefore, the amino acids immediately following the assembly strand are not essential for PTEX function. However, complementation with a version of EXP2 that lacks an additional 12 residues, removing the assembly strand, failed to rescue these phenotypes (Fig. 5c–f). These results demonstrate that the EXP2 assembly strand is critical to PTEX function, consistent with an essential role for it in docking the HSP101 unfoldase to the EXP2 membrane channel to facilitate translocation.

A strong, albeit lower resolution claw-shaped density extends from the end of each modelled PTEX150(S668–D832) shaft to the HSP101 middle domain above, terminating in a three-turn helix that rests on the midpoint of the middle domain. This helix forms a strong interaction with Y488 and Y491 of HSP101 in claws a–e (Extended Data Fig. 8d–e), but is not visible in claw f in the engaged state. Claw g appears to form an additional interaction with the N-terminal end of the middle domain of HSP101 protomer 1 (Extended Data Fig. 8d). The middle domain is known to have a critical role in regulating ATPase and unfoldase activities in related HSP100s[35,36], suggesting that this interaction is of high functional importance.

Two observed states suggest a translocation mechanism

In addition to the extended 195 Å engaged state of PTEX, we also observed a more compact 175 Å resetting state. Much of PTEX150(S668–D823) and EXP2 remain unchanged between the engaged and resetting states; a hinge-like swinging motion in the HSP101 hexamer accounts for the 20 Å difference in height. The transmembrane domain and B1–B3 helices of EXP2 exhibit C7 symmetry, and remain identical between the two states (Supplementary Video 6). The deviations from C7 symmetry in the B4–B5 helices and assembly domain are less pronounced in the resetting state (Extended Data Fig. 3b, c), probably owing to the more planar arrangement of HSP101 protomers. As in the engaged state, slight inter-protomer variations in the H2–H3 region of PTEX150(S668–D823) bridge the gap between EXP2 and HSP101, maintaining a continuous protected path for unfolded cargo proteins.

The spiral staircase of HSP101 tyrosine pore loops in the engaged state collapses into a planar C shape in the resetting state (Supplementary Video 7), with a freedom of movement possibly conferred by the gap between HSP101 protomers 1 and 6[30,31]. Starting at the interface between the NBD2 domains of HSP101 protomers 3 and 4, HSP101 protomers 4–6 swing downwards and outwards, creating a deep vertical cleft through the central pore of the hexamer.

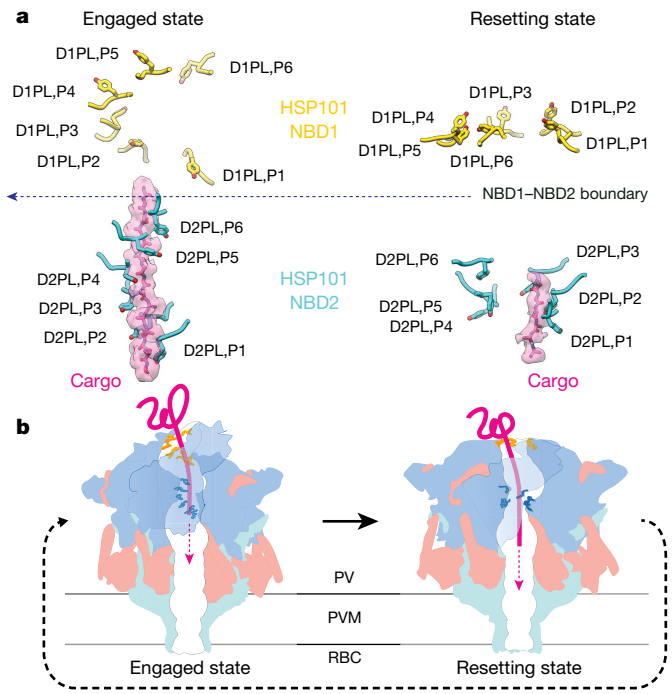


Fig. 6 | Mechanism of translocation. a, Side views of the HSP101 pore loops with the unfolded cargo peptide backbone models (dark pink) built into the cryo-EM densities (light pink). Vertical distances between pore loop tyrosines in consecutive loops are as follows. Engaged D1PL,P1–D1PL,P6: 9.41 Å, 8.61 Å, 1.40 Å, 3.34 Å and 2.28 Å, respectively. Engaged D2PL,P1–D2PL,P6: 6.52 Å, 6.28 Å, 6.38 Å, 6.96 Å and 6.12Å, respectively. Resetting D1PL,P1–D1PL,P6: 1.75 Å, −2.70 Å, −1.65 Å, −0.78 Å and 1.81 Å, respectively. Resetting D2PL,P1–D2PL,P6: 5.88 Å, 4.56 Å, −6.80 Å, 2.25 Å and 7.88 Å, respectively. **b**, Proposed stepwise feeding mechanism of translocation by PTEX. NBD1 and NBD2 pore loops and cargo are coloured as in **a**.

This movement pulls the NBD2 loops in protomers 4–6 away from the unfolded cargo (Extended Data Fig. 3d, e, h, i, Supplementary Videos 6, 7). A shorter (19 Å versus 45 Å), unfolded cargo density remains visible, bound to the NBD2 loops in protomers 1–3, whereas no peptide density is visible in protomers 4–6 (Fig. 6a, Extended Data Fig. 3h, i). Furthermore, the NBD1 domain of protomer 3 rotates outward, such that the R361 arginine finger remains within 5.2 Å of the ATPγS bound to NBD1 of protomer 4, and the nucleotide in NBD2 of protomer 4 shifts 7.5 Å away from the R859 arginine finger in protomer 3 (Extended Data Fig. 3f, g).

Discussion

We propose a PTEX-mediated mechanism of protein translocation via a cyclic process involving at least two discrete states (Fig. 6b, Supplementary Videos 6, 7), which we have captured by purifying PTEX complexes directly from *P. falciparum* that are actively translocating cargo. The pore loops in HSP101 NBD2 form two 'hands' that work together to thread the cargo protein through the central pore. NBD2 loops from HSP101 protomers 1–3 form the passive hand, located closest to the PTEX150(S668–D823)–EXP2 funnel, which remains fixed between states (Fig. 6a, b). NBD2 loops from HSP101 protomers 4–6 form the active hand, which moves along the channel axis (above the passive hand), grasping the unfolding peptide and feeding it through the passive hand. In the engaged state, all six NBD2 pore loops grip the unfolded peptide in the spiral staircase formation (Fig. 6a, b). As the HSP101 hexamer collapses into the resetting state, the active hand moves downwards, feeding the newly unfolded peptide through the passive hand, into the PTEX150(S668–D823)–EXP2 funnel below. The passive hand then grips the unfolded peptide, preventing it from slipping back towards the HSP101 apical entrance while the active hand swings outward, releasing the cargo (Fig. 6a, b). Finally, the active hand moves upwards to grasp the unfolding protein further upstream, transitioning back to the engaged state. This cyclic feeding mechanism threads the unfolded cargo protein through the translocon, across the PVM and into the cytosol of the host cell.

The states captured here may be two of several states in the processive phase of translocation. Additional states are likely to exist for cargo recognition. Although we did not observe PTEX-free HSP101 oligomers as has been suggested[37], we did observe additional, seemingly cargo-free PTEX complexes (Extended Data Fig. 9) that did not refine to better than 7 Å, suggesting that there is conformational heterogeneity in the absence of stabilizing cargo interactions. Cargo–PTEX interactions during cargo recognition may be transient, possibly explaining why we did not observe the N-terminal domains of HSP101, or other components that are potentially required for cargo recognition. Without these details, the mechanisms of cargo recognition and subsequent refolding after translocation remain unclear, although there is some evidence for the involvement of exported parasite chaperones[38] or co-opted host chaperonins[39] in these processes. Of note, on the basis of secondary structure prediction and PTEX150 truncation experiments[37], PTEX150 residues D838–F912 may occupy the claw (PTEX150 residues D838–E873) and three-turn helix (PTEX150 residues S884–F912) densities that remain unassigned in our structures.

Our work demonstrates the advantages of obtaining structures of challenging protein complexes in functionally relevant states by imaging samples purified directly from endogenous sources. Direct observation of the native PTEX core provides compelling evidence that this complex containing EXP2, PTEX150 and HSP101 is a bona fide translocon that is embedded in the PVM and serves as the gateway for the malaria parasite exportome. In addition to establishing the role of EXP2 as the membrane-spanning pore of PTEX and providing insight into the mechanism of this essential protein translocating machine, our structures reveal an interaction between the EXP2 assembly domain and the C-terminal domain of HSP101 that is indispensable for PTEX function. These atomic structures of PTEX provide potential targets for designing a new class of drugs to inhibit this essential gatekeeper of the malarial exportome.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, statements of data availability and associated accession codes are available at https://doi.org/10.1038/s41586-018-0469-4.

Received: 29 March 2018; Accepted: 19 July 2018;
Published online 27 August 2018.

1. World Health Organization. *World Malaria Report 2017*. (World Health Organization, Geneva, 2017).
2. Miller, L. H., Ackerman, H. C., Su, X. Z. & Wellems, T. E. Malaria biology and disease pathogenesis: insights for new treatments. *Nat. Med.* **19**, 156–167 (2013).
3. Cowman, A. F., Healer, J., Marapana, D. & Marsh, K. Malaria: biology and disease. *Cell* **167**, 610–624 (2016).
4. Spillman, N. J., Beck, J. R. & Goldberg, D. E. Protein export into malaria parasite-infected erythrocytes: mechanisms and functional consequences. *Annu. Rev. Biochem.* **84**, 813–841 (2015).
5. Boddey, J. A. & Cowman, A. F. *Plasmodium* nesting: remaking the erythrocyte from the inside out. *Annu. Rev. Microbiol.* **67**, 243–269 (2013).
6. Przyborski, J. M., Nyboer, B. & Lanzer, M. Ticket to ride: export of proteins to the *Plasmodium falciparum*-infected erythrocyte. *Mol. Microbiol.* **101**, 1–11 (2016).
7. de Koning-Ward, T. F., Dixon, M. W., Tilley, L. & Gilson, P. R. *Plasmodium* species: master renovators of their host cells. *Nat. Rev. Microbiol.* **14**, 494–507 (2016).
8. Marti, M., Good, R. T., Rug, M., Knuepfer, E. & Cowman, A. F. Targeting malaria virulence and remodeling proteins to the host erythrocyte. *Science* **306**, 1930–1933 (2004).
9. Hiller, N. L. et al. A host-targeting signal in virulence proteins reveals a secretome in malarial infection. *Science* **306**, 1934–1937 (2004).
10. Heiber, A. et al. Identification of new PNEPs indicates a substantial non-PEXEL exportome and underpins common features in *Plasmodium falciparum* protein export. *PLoS Pathog.* **9**, e1003546 (2013).
11. Lingelbach, K. & Joiner, K. A. The parasitophorous vacuole membrane surrounding *Plasmodium* and *Toxoplasma*: an unusual compartment in infected cells. *J. Cell Sci.* **111**, 1467–1475 (1998).

12. Ansorge, I., Benting, J., Bhakdi, S. & Lingelbach, K. Protein sorting in *Plasmodium falciparum*-infected red blood cells permeabilized with the pore-forming protein streptolysin O. *Biochem. J.* **315**, 307–314 (1996).
13. Gehde, N. et al. Protein unfolding is an essential requirement for transport across the parasitophorous vacuolar membrane of *Plasmodium falciparum*. *Mol. Microbiol.* **71**, 613–628 (2009).
14. de Koning-Ward, T. F. et al. A newly discovered protein export machine in malaria parasites. *Nature* **459**, 945–949 (2009).
15. Beck, J. R., Muralidharan, V., Oksman, A. & Goldberg, D. E. PTEX component HSP101 mediates export of diverse malaria effectors into host erythrocytes. *Nature* **511**, 592–595 (2014).
16. Elsworth, B. et al. PTEX is an essential nexus for protein export in malaria parasites. *Nature* **511**, 587–591 (2014).
17. Bullen, H. E. et al. Biosynthesis, localization, and macromolecular arrangement of the *Plasmodium falciparum* translocon of exported proteins (PTEX). *J. Biol. Chem.* **287**, 7871–7884 (2012).
18. Johnson, D. et al. Characterization of membrane proteins exported from *Plasmodium falciparum* into the host erythrocyte. *Parasitology* **109**, 1–9 (1994).
19. Gold, D. A. et al. The *Toxoplasma* dense granule proteins GRA17 and GRA23 mediate the movement of small molecules between the host and the parasitophorous vacuole. *Cell Host Microbe* **17**, 642–652 (2015).
20. Garten, M. et al. R. EXP2 is a nutrient-permeable channel in the vacuolar membrane of *Plasmodium* and is essential for protein export via PTEX. *Nat. Microbiol.* https://doi.org/10.1038/s41564-018-0222-7 (2018).
21. Chisholm, S. A. et al. The malaria PTEX component PTEX88 interacts most closely with HSP101 at the host-parasite interface. *FEBS J.* **285**, 2037–2055 (2018).
22. Holm, L. & Laakso, L. M. Dali server update. *Nucleic Acids Res.* **44**, W351–W355 (2016).
23. Gibrat, J. F., Madej, T. & Bryant, S. H. Surprising similarities in structure comparison. *Curr. Opin. Struct. Biol.* **6**, 377–385 (1996).
24. Wiederstein, M., Gruber, M., Frank, K., Melo, F. & Sippl, M. J. Structure-based characterization of multiprotein complexes. *Structure* **22**, 1063–1070 (2014).
25. Krissinel, E. & Henrick, K. Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. *Acta Crystallogr. D Biol. Crystallogr.* **60**, 2256–2268 (2004).
26. Dosztanyi, Z., Csizmok, V., Tompa, P. & Simon, I. IUPred: web server for the prediction of intrinsically unstructured regions of proteins based on estimated energy content. *Bioinformatics* **21**, 3433–3434 (2005).
27. Dosztanyi, Z., Csizmok, V., Tompa, P. & Simon, I. The pairwise energy content estimated from amino acid composition discriminates between folded and intrinsically unstructured proteins. *J. Mol. Biol.* **347**, 827–839 (2005).
28. Schirmer, E. C., Glover, J. R., Singer, M. A. & Lindquist, S. HSP100/Clp proteins: A common mechanism explains diverse functions. *Trends Biochem. Sci.* **21**, 289–296 (1996).
29. Hanson, P. I. & Whiteheart, S. W. AAA+ proteins: have engine, will work. *Nat. Rev. Mol. Cell Biol.* **6**, 519–529 (2005).
30. Deville, C. et al. Structural pathway of regulated substrate transfer and threading through an Hsp100 disaggregase. *Sci. Adv.* **3**, e1701726 (2017).
31. Gates, S. N. et al. Ratchet-like polypeptide translocation mechanism of the AAA+ disaggregase Hsp104. *Science* **357**, 273–279 (2017).
32. AhYoung. A. P., Koehl, A., Cascio, D. & Egea, P. F. Structural mapping of the ClpB ATPases of *Plasmodium falciparum*: targeting protein folding and secretion for antimalarial drug design. *Protein Sci.* **24**, 1508–1520 (2015).
33. Puchades, C. et al. Structure of the mitochondrial inner membrane AAA+ protease YME1 gives insight into substrate processing. *Science* **358**, (2017).
34. Yokom, A. L. et al. Spiral architecture of the Hsp104 disaggregase reveals the basis for polypeptide translocation. *Nat. Struct. Mol. Biol.* **23**, 830–837 (2016).
35. Seyffer, F. et al. Hsp70 proteins bind Hsp100 regulatory M domains to activate AAA+ disaggregase at aggregate surfaces. *Nat. Struct. Mol. Biol.* **19**, 1347–1355 (2012).
36. Lipinska, N. et al. Disruption of ionic interactions between the nucleotide binding domain 1 (NBD1) and middle (M) domain in Hsp100 disaggregase unleashes toxic hyperactivity and partial independence from Hsp70. *J. Biol. Chem.* **288**, 2857–2869 (2013).
37. Elsworth, B. et al. Proteomic analysis reveals novel proteins associated with the *Plasmodium* protein exporter PTEX and a loss of complex stability upon truncation of the core PTEX component, PTEX150. *Cell. Microbiol.* **18**, 1551–1569 (2016).
38. Kulzer, S. et al. *Plasmodium falciparum*-encoded exported hsp70–hsp40 chaperonen–co-chaperone complexes within the host erythrocyte. *Cell. Microbiol.* **14**, 1784–1795 (2012).
39. Batinovic, S. et al. An exported protein-interacting complex involved in the trafficking of virulence determinants in *Plasmodium*-infected erythrocytes. *Nat. Commun.* **8**, 16044 (2017).
40. Smart, O. S., Neduvelil, J. G., Wang, X., Wallace, B. A. & Sansom, M. S. P. HOLE: A program for the analysis of the pore dimensions of ion channel structural models. *J. Mol. Graph.* **14**, 354–360 (1996).

Acknowledgements This research was supported in part by grants from National Institutes of Health (R21AI125983 to P.F.E., R01GM071940/AI094386/DE025567 to Z.H.Z. and K99/R00 HL133453 to J.R.B.). P.F.E. is the Alexander and Renee Kolin Endowed Chair in Molecular Biology and Biophysics. C.-M.H. acknowledges funding from the Ruth L. Kirschstein National Research Service Award (AI007323). We thank the UCLA Proteome Research Center for assistance in mass spectrometry and acknowledge the use of instruments in the Electron Imaging Center for Nanomachines supported by UCLA and grants from NIH (S10RR23057, S10OD018111 and U24GM116792) and NSF (DBI-1338135 and DMR-1548924). We thank A. W. P. Fitzpatrick for input on cryo-EM aspects of the project and J. Su for helping with Fig. 1a.

Reviewer information *Nature* thanks A. Cowman, P. da Fonseca and W. Hol for their contribution to the peer review of this work.

Author contributions C.-M.H., P.F.E. and Z.H.Z. initiated the project; J.R.B. generated *P. falciparum* lines, collected *P. falciparum*, performed complementation experiments and helped to write the paper; C.-M.H. purified the sample, screened purified samples by negative stain, prepared cryo-EM grids, acquired and processed the cryo-EM data, interpreted the structures and wrote the paper; M.L. built and refined the atomic models and helped to interpret the structures; Y.C. helped with sample freezing; Z.H.Z. supervised the cryo-EM aspects of the project, interpreted the structures and wrote the paper; P.F.E. supervised biochemical aspects of the project and helped to interpret the structures; D.E.G. supervised parasitology aspects of the project. D.E.G., M.L. and P.F.E. helped to edit the paper. Queries regarding cryo-EM and atomic modelling should be addressed to Z.H.Z. and queries on biochemistry should be addressed to P.F.E.

Competing interests The authors declare no competing interests.

Additional information

Extended data is available for this paper at https://doi.org/10.1038/s41586-018-0469-4.
Supplementary information is available for this paper at https://doi.org/10.1038/s41586-018-0469-4.
Reprints and permissions information is available at http://www.nature.com/reprints.
Correspondence and requests for materials should be addressed to P.F.E. or Z.H.Z.

METHODS

No statistical methods were used to predetermine sample size. The experiments were not randomized. The investigators were not blinded to allocation during experiments and outcome assessment.

Cells. *P. falciparum* strain NF54^{attB} was obtained from the Fidock laboratory where it was generated[41] and was used exclusively in the study. De-identified, IRB-exempt expired RBCs were obtained from the blood bank at the St. Louis Children's Hospital. PCR amplified regions from the NF54^{attB} genome were found to match the genome sequence for 3D7, a sub clone of NF54. The presence of the cg6 localized attB sequence was verified by successful Bxb1-mediated integration at that site. Cell lines were not tested for mycoplasma contamination.

***P. falciparum* culture and genetic modification for PTEX purification.** *P. falciparum* culture was performed as described[42] with the exception that RPMI was supplemented with 0.5% Albumax I. All plasmid construction was carried out by infusion cloning (Clontech) unless otherwise noted. Integration of a 3×Flag fusion at the endogenous HSP101 C terminus was accomplished with CRISPR–Cas9 editing. A Cas9 target site was chosen just upstream of the *hsp101* stop codon (TAATAGTAAAGCTAAAAACT) and the guide RNA seed sequence was synthesized as a sense and anti-sense primer pair (sense shown) 5′-TAAGTATATAATATTTAATAGTAAAGCTAAAAACTGTTTTAGAGCTAGAA-3′, annealed and inserted into the BtgZI site of the plasmid pAIO[43], resulting in the plasmid pAIO-HSP101-CT-gRNA1. A 5′-homology flank (up to but not including the stop codon) was amplified from *P. falciparum* NF54^{attB} genomic DNA using primers 5′-GACGCGAGGAAAATTAGCATGCATCCTTAAGGAGATTCTGGTATGCCACTTGGTTC-3′ and 5′-CTGCACCTGGCCTAGGGGTCTTAGATAAGTTTATAACTAAGTTTTTAGCTTTACTATT-3′, incorporating a synonymous shield mutation in the protospacer adjustment motif of the gRNA target site within the *hsp101* coding sequence. A 3′-homology flank (beginning 3 bp downstream of the stop codon) was amplified using primers 5′-CACTATAGAACTCGAGAATTACGCATATATATATATATATATATATATAACATGGGTTG-3′ and 5′-GAACCAAGTGGCATACCAGAATCTCCTTAAGGATGCATGCTAATTTTCCTCGCGTC-3′. The flank amplicons were assembled in a second PCR reaction using primers 5′-CACTATAGAACTCGAGAATTACGCATATATATATATATATATATATATAACATGGGTTG-3′ and 5′-CTGCACCTGGCCTAGGGGTCTTAGATAAGTTTATAACTAAGTTTTTAGCTTTACTATT-3′ and inserted between XhoI and AvrII in pPM2GT[42]. The GFP tag between AvrII and EagI in this vector was then replaced with sequence encoding a 3×Flag tag using the primer 5′-CTTAGTTATAAACTTATCTAAGACCCCTAGGGACTACAAGGACGACGACGACAAGGATTATAAAGATGATGATGATAAAGATTATAAAGATGATGATGATAAATGACGGCCGCGTCGAGTTATATAATATATTTATG-3′ and a QuikChange Lightning Multi Site-Directed Mutagenesis kit (Agilent), resulting in the plasmid pPM2GT-HSP101-3×Flag. This plasmid was linearized at the AflII site between the 3′ and 5′-homology flanks and co-transfected with pAIO-HSP101-CT-gRNA1 into *P. falciparum* NF54^{attB} parasites[41]. Selection with 10 nM WR99210 was applied 24 h after transfection. Once *P. falciparum* returned from selection, integration at the intended site was confirmed by PCR with primers 5′-CGAAAACTTTTATGGTATTAATATAACAG-3′ and 5′-CCTTGTCGTCGTCGTCCTTG-3′ and a clonal line was isolated by limiting dilution.

For PTEX purification, *P. falciparum* expressing HSP101-3×Flag were synchronized by serial treatment with 5% w/v D-sorbitol and then expanded while shaking to increase singlet invasion events and maintain synchrony. For each preparation, ~2×10^{10} erythrocytes infected with *P. falciparum* were collected at the ring stage (typically ~500 ml 2% haematocrit culture at ~20% parasitaemia). Erythrocytes were lysed in 10× pellet volume of cold phosphate-buffered saline (PBS) containing 0.0125% saponin (Sigma, sapogenin content ≥ 10%) and EDTA-free protease inhibitor cocktail (Roche or Pierce). Released *P. falciparum* were washed in cold PBS containing EDTA-free protease inhibitor cocktail and washed cell pellets were frozen in liquid nitrogen and stored at −80 °C.

Affinity purification of PTEX core complex from *P. falciparum* pellets. Frozen *P. falciparum* pellets were resuspended in lysis buffer (25 mM HEPES pH 7.4, 10 mM $MgCl_2$, 150 mM KCl, 10% glycerol) and homogenized using a glass Dounce tissue homogenizer. The membrane fraction was isolated from the homogenized lysate by centrifugation at 100,000*g* for 1 h. The membrane pellet was solubilized in solubilization buffer (25 mM HEPES pH 7.4, 10 mM $MgCl_2$, 150 mM KCl, 10% glycerol, 0.4% triton X-100) and the solubilized membranes were then applied to anti-Flag M2 affinity gel resin (Sigma). The resin was washed extensively in wash buffer (25 mM HEPES pH 7.4, 10 mM $MgCl_2$, 150 mM KCl, 10% glycerol, 0.015% triton X-100), after which the protein was eluted from the affinity resin with elution buffer (25 mM HEPES pH 7.4, 10 mM $MgCl_2$, 150 mM KCl, 2 mM ATPγS, 0.015% triton X-100, 500 μg/ml Flag peptide).

The presence and relative abundance of the three PTEX core components were verified by silver stained SDS–PAGE and tryptic digest liquid chromatography–mass spectrometry (Extended Data Fig. 1e, f). The extremely low yields achievable when purifying PTEX directly from *P. falciparum* prohibited the conventional approach of evaluating sample quality by size exclusion chromatography. Thus, during the iterative process of screening for optimal purification conditions, sample quality was assessed by negative stain (uranyl acetate) transmission electron microscopy in an FEI TF20 microscope equipped with a TVIPS 16 mega-pixel CCD camera. In brief, small datasets of ~100,000 particles were collected and 2D class averages were generated in RELION[44,45] to assess the presence of sufficient numbers of intact PTEX particles yielding 'good' class averages exhibiting distinct features. For example, C7 symmetry could be recognized in top views, and the characteristic Clp/HSP100 layers were visible in side views (Extended Data Fig. 6a–c).

Cryo-electron microscopy. Three-microlitre aliquots of purified PTEX core complex were applied to glow-discharged lacey carbon grids with a supporting ultrathin carbon film (Ted Pella). Grids were then blotted with filter paper and vitrified in liquid ethane using an FEI Vitrobot Mark IV or a home-made manual plunger. Cryo-EM grids were screened in an FEI Tecnai TF20 transmission electron microscope while optimizing freezing conditions.

Higher resolution cryo-EM images were collected on a Gatan K2-Summit direct electron detector in counting mode on an FEI Titan Krios at 300 kV equipped with a Gatan Quantum energy filter set at a 20 eV slit width. Fifty frames were recorded for each movie at a pixel size of 1.04 Å at the specimen scale, with a 200-ms exposure time and an average dose rate of 1.2 electrons per $Å^2$ per frame, resulting in a total dose of 60 electrons per $Å^2$ per movie. The final dataset consists of a total of 25,000 movies recorded in four separate sessions.

Image processing and 3D reconstruction. Frames in each movie were aligned, gain reference-corrected and dose-weighted to generate a micrograph using MotionCor2[46]. Aligned and un-dose-weighted micrographs were also generated and used for contrast transfer function (CTF) estimation using CTFFIND4[47] and PTEX particle picking by hand and using Gautomatch (https://www.mrc-lmb.cam.ac.uk/kzhang/Gautomatch/Gautomatch_Brief_Manual.pdf).

In total, 1,508,462 particles were extracted from 19,752 micrographs and initially binned by a factor of 2. After two rounds of reference-free two-dimensional (2D) classification in RELION, 422,713 particles were selected as 'good' particles from distinct 2D class averages representing different views of the PTEX core complex. These particles were then used in a one-class ab initio reconstruction followed by homogeneous refinement in CryoSPARC[48], yielding a 4.8 Å ab initio 3D map.

The original 422,713 'good' particles were then aligned in a 3D refinement in RELION using the 4.8 Å CryoSPARC map as an initial reference. All subsequent image-processing steps were performed using RELION. After this refinement, the particles were unbinned, their centres recalculated and used to re-extract particles from the original micrographs without binning. The newly extracted, unbinned particles were then aligned with a second 3D refinement yielding a ~4.5 Å reconstruction.

An exhaustive, iterative search of classification and refinement conditions was used to sort out different conformations and further improve resolution (Extended Data Fig. 9). In brief, upon further sorting using 3D-classification without alignment, we identified two homogenous particle subsets corresponding to the engaged and resetting states (Extended Data Fig. 9). Particles in the two subsets were refined separately, yielding full maps with overall resolutions of 4.16 Å and 4.23 Å, respectively.

Focused 3D classification without alignment followed by focused refinement was used to further improve the resolution of mobile regions of the structure in both states. C7 symmetry was applied in the focused 3D classification and refinement steps of the heptameric halves, comprising EXP2 and PTEX150, yielding a 3.4 Å engaged state map and a 3.5 Å resetting state map (Extended Data Figs. 3, 9). The same procedure, except with C1 symmetry, was applied to the hexameric half of the engaged state, yielding a 4.09 Å map (Extended Data Figs. 3, 9). This last step was also applied to the hexameric half of the resetting state, but did not yield improvements in resolution. Further efforts of focused 3D classification and refinement of individual HSP101 protomers, individual claws, and HSP101 N-terminal domain densities in the two states did not ultimately yield improvements in resolution in either state.

Model building and refinement. Map interpretation was performed with UCSF Chimera[49] and COOT[50]. *P. falciparum* protein sequences were obtained from the National Center for Biotechnology Information (NCBI)[51] and the PlasmoDB[52] protein databases. PHYRE2[53] secondary structure predictions were used as an aid for initial manual sequence registration. Models for a single monomer of HSP101, PTEX150, and EXP2 in the engaged state were all built de novo. This first model for each protein monomer was then placed into the density maps of other protomers to aid de novo modelling of subsequent protomers. Individual protomers in the complex were then manually remodelled to ensure a close fit between densities and models. The same process was repeated for the resetting state. Manual refinement targeting protein geometry alone was done primarily along the periphery and flexible regions of the complex (for example, the M domains of HSP101). Whereas

their densities and backbone traces were visible, we were unable to model the claw with its connected three-turn helix, nor one of the 12 M domain loops in the resetting state (Fig. 5g, h). The three-turn helix displayed a few bulky side chains interacting with the M domain of HSP101; however, the lack of backbone connection to our atomic model of the complex and the limited visibility of smaller side chains in this region made sequence assignment challenging.

Manual refinement targeting both protein geometry and fit with the density map was used primarily in the core regions where resolution was higher and noise was minimal. Rotamers were fitted manually in COOT and improved using the 'back-rub rotamers' setting. The resulting models for the complexes were subjected to the phenix.real_space_refine program in PHENIX[54]. Following this step, Molprobity[55] reported less than ideal clash scores and map-to-model cross-correlation. To improve the geometry and fit, manual adjustments were made to protein geometry and density map fit, with the additional step of using Molprobity[55] clash dots and sphere-refinement in COOT.

The complex was then broken into three portions: (1) symmetric regions of EXP2 and PTEX150, (2) HSP101, and (3) the full PTEX complex. These model segments were fed back to phenix.geometry_minimization in PHENIX and then to phenix.real_space_refine using simulated annealing and global minimization applying Emsley's Ramachandran restraints[50]. Following another round of manual checks and improvements, all models were subjected to phenix.real_space_refine with default settings one last time.

All figures and videos were prepared with UCSF Chimera, Pymol (https://www.pymol.org), and Resmap[56]. Molprobity was used to validate the stereochemistry of the final models.

Genetic complementation. For expression of a complementing second copy of truncated EXP2, the *exp2* coding sequence up to codon position 221 was amplified with primers 5′-CGAATAAACACGATTTTTTCTCGAGATGAAAGTCAGTT ATATATTTTCCTTTTTTTTGTTATTCTTCG-3′ and 5′-AATCAACTT TTGTTCGCTAGCTTTCTTTGATTCCATAGATTTCAATTTCTCTTCC-3′ and inserted into the plasmid pyEOE-attP-EXP2-3×Myc[20] between XhoI and NheI, resulting in the plasmid pyEOE-attP-EXP2Δ222-287-3×Myc. This plasmid was co-transfected with pINT[41] into EXP2^{apt}::HSP101-3×Flag conditional knockdown parasites[20] at the mature schizont stage using a Nucleofector 2b and Basic Parasite Nucleofector kit 2 (Lonza). Selection with 2 μM DSM1[57] was applied 24 h post-transfection (in addition to 2.5 μg/ml blasticidin S and 1 μM anhydrotetracycline for maintenance of endogenous EXP2 translational control by the aptamer system) to facilitate integration into the attP site engineered in the benign *cg6* locus through integrase mediated attB × attP recombination. Following return from selection, *P. falciparum* were cloned by limiting dilution, and expression of EXP2(Δ222–287)–3×Myc was confirmed by western blot.

***P. falciparum* growth assays.** EXP2^{apt}::EXP2(Δ222–287) *P. falciparum* were extensively washed to remove aTc and plated with or without 1 μM aTc in triplicate at an initial parasitaemia of 1%. The medium was changed every 48 h and 1:1 subculture was performed every other day beginning on day 4 to avoid culture overgrowth. Parasitaemia (percentage of total red blood cells (RBCs) infected) was measured every 24 h by flow cytometry on a FACSCanto (BD Biosciences) by nucleic acid staining of cultured RBCs with PBS containing 0.8 μg/ml acridine orange. Cumulative parasitaemias were back calculated based on the subculture schedule and data were fitted to an exponential growth equation to determine rate constants using Prism (Graphpad).

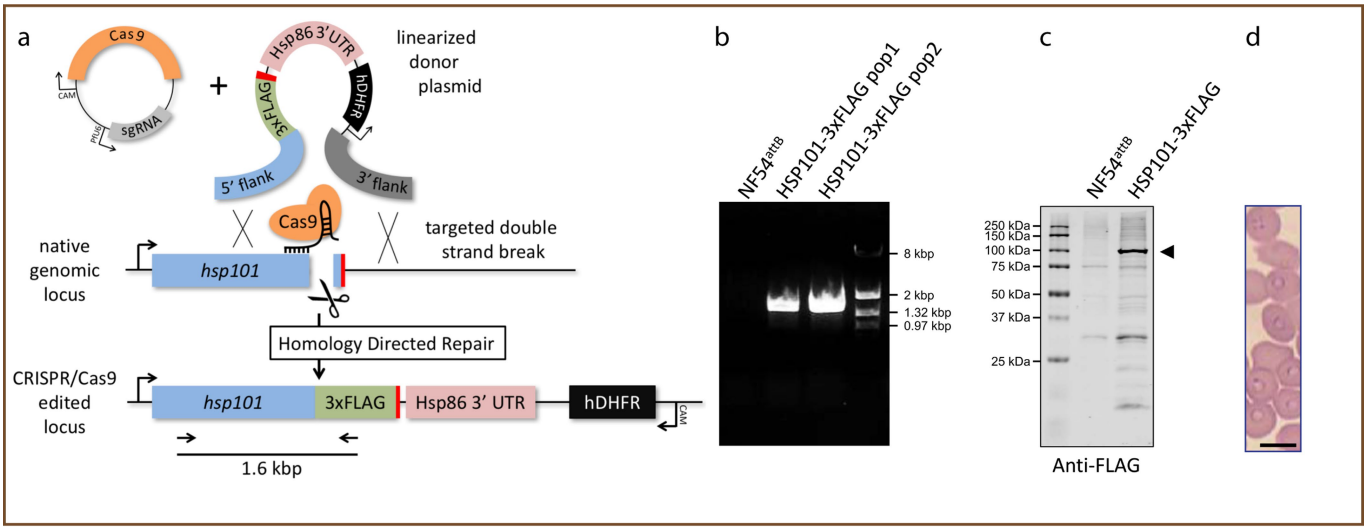
Quantification of protein export. For evaluation of protein export by immunofluorescence assay (IFA), mature schizonts were purified on a magnetic column and allowed to invade fresh, uninfected RBCs with shaking for 3 h before treatment with 5% w/v D-sorbitol to destroy unruptured schizonts. Pulse-invaded cells were plated with or without 1 μM aTc and allowed to develop 24 h post-invasion. Thin smears of infected RBCs were briefly air dried and immediately fixed in ice-cold acetone for 2 min. After fixation, samples were blocked for 30 min in PBS + 3% BSA followed by incubation for 1 h with primary antibody solutions containing mouse anti-Flag M2 monoclonal antibody (detecting HSP101–3×Flag to mark the PVM) and rabbit anti-SBP1. After washing, secondary antibody incubation was carried out for 1 h with Alexa Fluor anti-mouse 488 and anti-rabbit 594 IgG antibodies (Life Technologies), each diluted 1:2,000. After final washing, coverslips were mounted over each sample using Pro-long antifade Gold with DAPI (Life Technologies). Images were collected with an ORCA-ER CCD camera (Hamamatsu) using AxioVision software on an Axio Imager.M1 microscope (Zeiss) with a 100× oil-immersion objective using the same exposure times for each image (300 ms for SBP1-594, 150 ms for Flag-488). Ten images were acquired for each condition using the DAPI channel for field selection to avoid bias. Images were then analysed using Volocity 6.3 (PerkinElmer). The border of each single-infected erythrocyte was traced using the DIC channel as a guide to define a region of interest (ROI). The PVM was marked using the 'find objects' measurement tool for the HSP101–3×Flag–488 channel (automatic threshold setting with threshold offset set to −30% and minimum object size set to 0.5 μm^2). Individual Maurer's clefts were identified using the 'find spots' measurement tool for the SBP1-594 channel (offset minimum spot intensity set to 40% and brightest spot within radius set to 0.5 μm). All spots within the PVM object boundary were then removed using the 'subtract' measurement tool and the number and fluorescent intensity of the remaining spots in each ROI were collected. Data were pooled from two independent experiments and plotted with Prism.

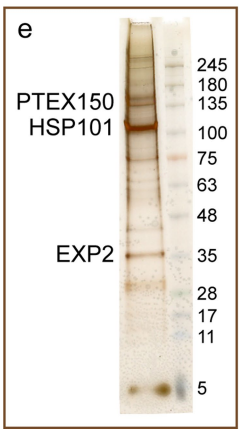
Antibodies. The following primary antibodies were used for IFA and western blot: mouse anti-Flag monoclonal antibody clone M2 (Sigma) (IFA, 1:500; western blot, 1:500); rabbit polyclonal anti-SBP1[58] (IFA, 1:500); mouse anti-cMYC monoclonal antibody 9E10 (ThermoFisher) (western blot, 1:300).

Reporting summary. Further information on research design is available in the Nature Research Reporting Summary linked to this paper.

Data availability. The atomic models have been deposited to the Protein Data Bank under accession numbers 6E10 and 6E11 and the cryo-EM density maps have been deposited in the Electron Microscopy Data Bank under accession numbers EMD-8951 and EMD-8952.

41. Adjalley, S. H. et al. Quantitative assessment of *Plasmodium falciparum* sexual development reveals potent transmission-blocking activity by methylene blue. *Proc. Natl Acad. Sci. USA* **108**, E1214–E1223 (2011).
42. Klemba, M., Beatty, W., Gluzman, I. & Goldberg, D. E. Trafficking of plasmepsin II to the food vacuole of the malaria parasite *Plasmodium falciparum*. *J. Cell Biol.* **164**, 47–56 (2004).
43. Spillman, N. J., Beck, J. R., Ganesan, S. M., Niles, J. C. & Goldberg, D. E. The chaperonin TRiC forms an oligomeric complex in the malaria parasite cytosol. *Cell Microbiol.* **19**, e12719 (2017).
44. Scheres, S. H. W. A Bayesian view on cryo-EM structure determination. *J. Mol. Biol.* **415**, 406–418 (2012).
45. Scheres, S. H. W. RELION: implementation of a Bayesian approach to cryo-EM structure determination. *J. Struct. Biol.* **180**, 519–530 (2012).
46. Zheng, S. Q. et al. MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy. *Nat. Methods* **14**, 331–332 (2017).
47. Rohou, A. & Grigorieff, N. CTFFIND4: fast and accurate defocus estimation from electron micrographs. *J. Struct. Biol.* **192**, 216–221 (2015).
48. Punjani, A., Rubinstein, J. L., Fleet, D. J. & Brubaker, M. A. cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. *Nat. Methods* **14**, 290–296 (2015).
49. Pettersen, E. F. et al. UCSF chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004).
50. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. D* **66**, 486–501 (2010).
51. NCBI Resource Coordinators. Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res.* **44**, D7–D19 (2016).
52. Aurrecoechea, C. et al. PlasmoDB: a functional genomic database for malaria parasites. *Nucleic Acids Res.* **37**, D539–D543 (2009).
53. Kelley, L. A., Mezulis, S., Yates, C. M., Wass, M. N. & Sternberg, M. J. E. The Phyre2 web portal for protein modeling, prediction and analysis. *Nat. Protoc.* **10**, 845–858 (2015).
54. Adams, P. D. et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D* **66**, 213–221 (2010).
55. Chen, V. B. et al. MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. D* **66**, 12–21 (2010).
56. Kucukelbir, A., Sigworth, F. J. & Tagare, H. D. Quantifying the local resolution of cryo-EMEM density maps. *Nat. Methods* **11**, 63–65 (2014).
57. Ganesan, S. M. et al. Yeast dihydroorotate dehydrogenase as a new selectable marker for *Plasmodium falciparum* transfection. *Mol. Biochem. Parasitol.* **177**, 29–34 (2011).
58. Blisnick, T. et al. Pfsbp 1, a Maurer's cleft *Plasmodium falciparum* protein, is associated with the erythrocyte skeleton. *Mol. Biochem. Parasitol.* **111**, 107–121 (2000).
59. Puchades, C. et al. Structure of the mitochondrial inner membrane AAA+ protease YME1 gives insight into substrate processing. *Science* **358**, eaao0464 (2017).

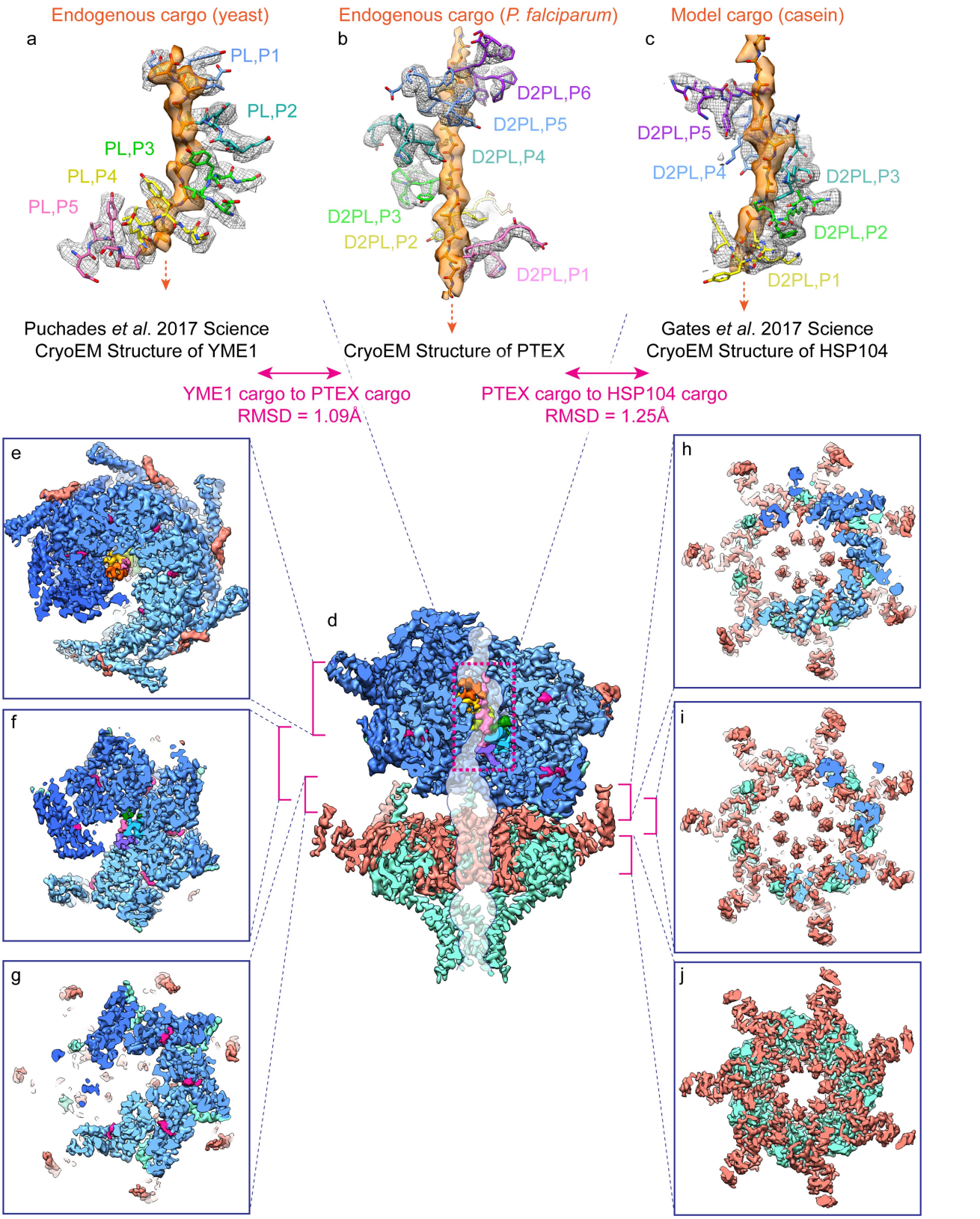




f

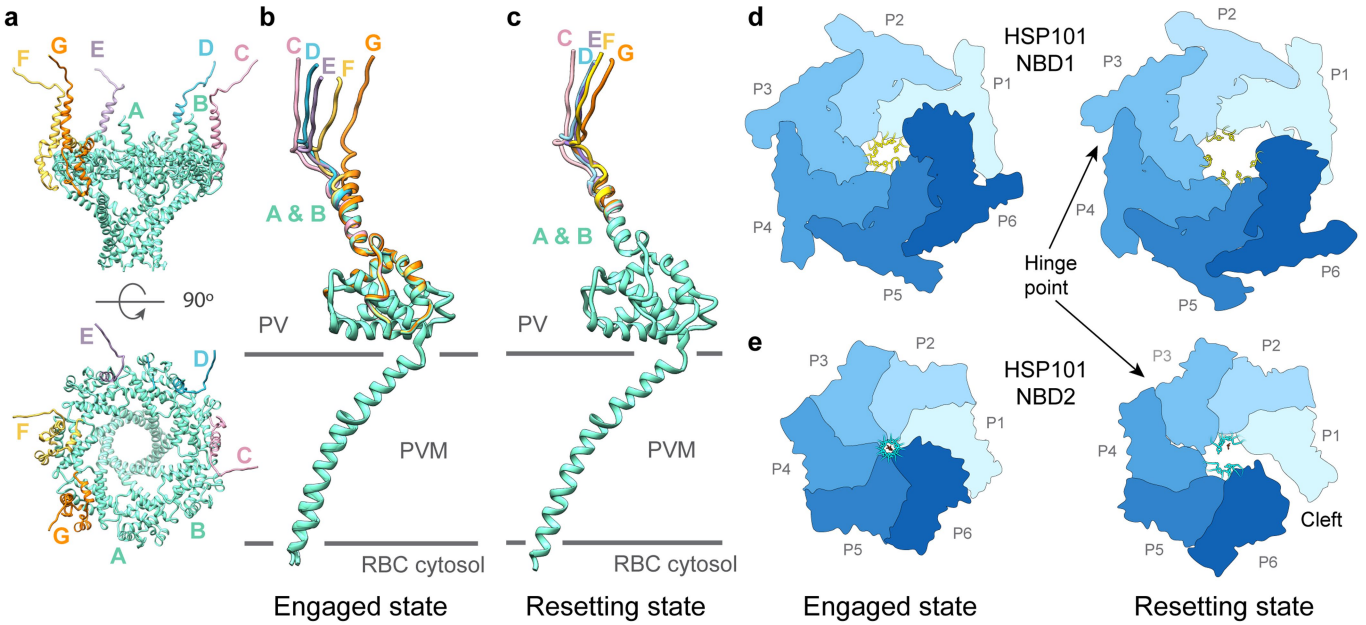
	Accession	Seq coverage	Seq count	Spec count	NSAF	EMPAI	Description
1	PF3D7_1116800	55.5	72	228	0.0750967	2.5892193	product=heat shock protein 101 (HSP101)
2	PF3D7_0917900	48.3	44	101	0.04622617	2.040885	product=heat shock protein 70 (HSP70-2)
3	PF3D7_1436300	58	58	135	0.04056941	2.8018937	product=translocon component PTEX150 (PTEX150)
4	PF3D7_1471100	40.8	13	34	0.03535177	1.5585859	product=exported protein 2 (EXP2)
5	PF3D7_1006200	59.8	7	7	0.019522185	2.9627802	product=DNA/RNA-binding protein Alba 3 (ALBA3)

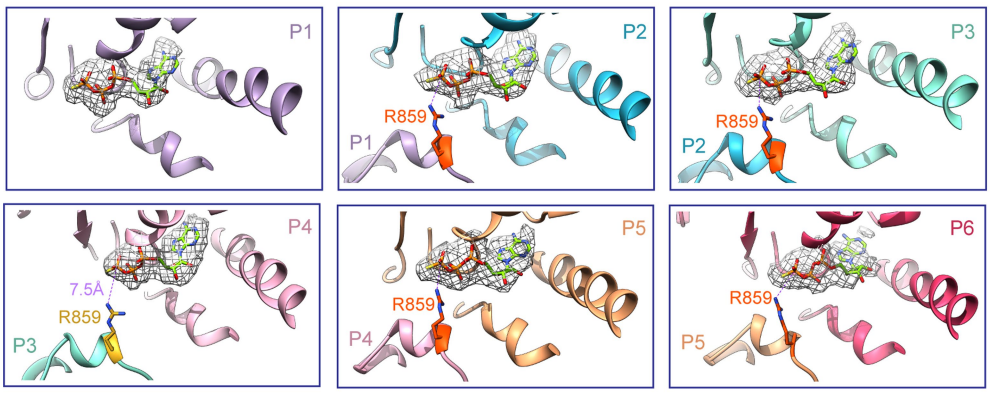
Extended Data Fig. 1 | Generation of HSP101–3×Flag *P. falciparum* and analysis of purified PTEX. a, Schematic showing strategy for endogenous tagging of *P. falciparum hsp101* with 3×Flag using CRISPR–Cas9 editing. Diagnostic PCR primers and expected amplicon following successful integration are shown. sgRNA, single-guide RNA; UTR, untranslated region; CAM, calmodulin promoter; PfU6, *P. falciparum* U6 promoter; hDHFR, human dihydrofolate reductase. **b**, Diagnostic PCR with genomic DNA template from NF54^{attB} parent or two independent populations of HSP101–3×Flag *P. falciparum*. The experiment was performed once. **c**, Western blot of NF54^{attB} and HSP101–3×Flag *P. falciparum* probed with mouse Flag M2 antibody (Sigma) and goat anti-mouse IRDye 680 secondary antibody (Li-Cor). Arrowhead indicates full-length HSP101–3×Flag (predicted molecular weight 102.9 kDa after signal peptide cleavage). Data represent two independent experiments. **d**, Giemsa staining of parasite-infected human erythrocytes from which PTEX was purified. Scale bar, 5 μm. For source data, see Supplementary Fig. 3. **e**, Silver-stained SDS–PAGE of the Flag-purified PTEX sample. Identities of the bands labelled EXP2, PTEX150 and HSP101 were confirmed by tryptic digest liquid chromatography–mass spectrometry (LC–MS). **f**, Tryptic digest LC–MS analysis of the Flag-purified PTEX sample. The PTEX core components are among the five most abundant species detected in the purified sample. For gel and blot source data, see Supplementary Fig. 1.

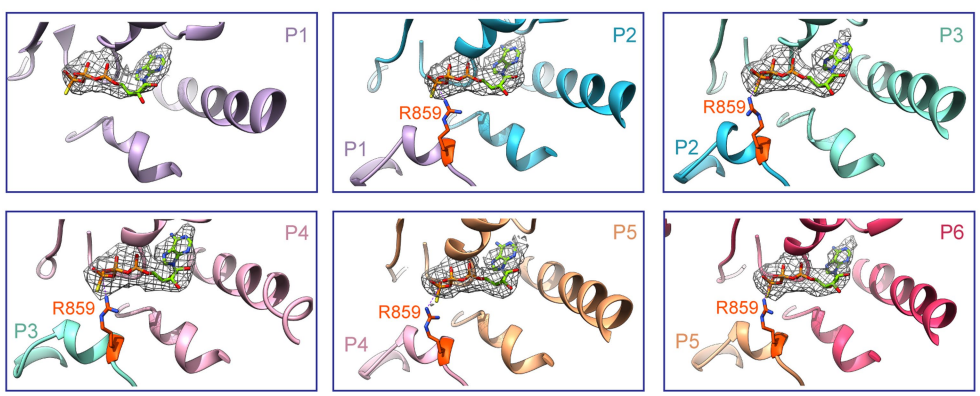


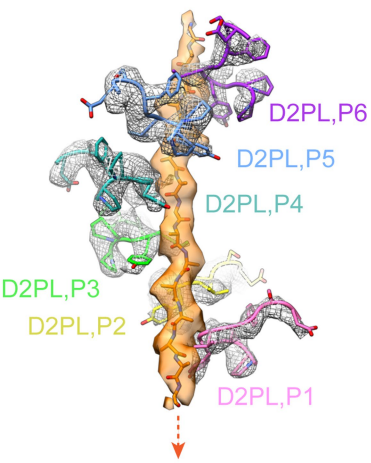
Extended Data Fig. 2 | See next page for caption.

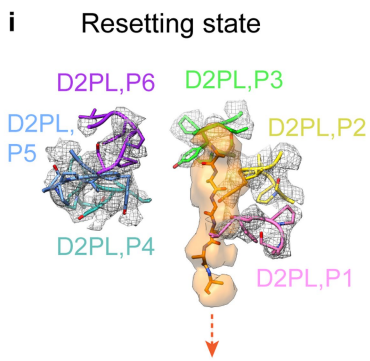
Extended Data Fig. 2 | Detailed views of the PTEX protein-conducting channel and symmetry mismatch in the engaged state. a–c, Cryo-EM densities and atomic models of cargo and pore loops from the near-atomic resolution structures of Clp/HSP100 ATPases YME1[59] (**a**), PTEX HSP101 (**b**) and HSP104[31] (**c**). Tyrosine sidechain densities are clearly visibly intercalating with the cargo densities. The modelled engaged state PTEX cargo has a calculated r.m.s.d. of 1.09 Å and 1.25 Å to the published YME1 and HSP104 cargo models, respectively. Pore loops are labelled by NBD and protomer (for example, D2PL,P1 is NBD2 pore loop, protomer 1). **d**, Side view of the bisected engaged state PTEX cryo-EM map. The protein-conducting channel, calculated using HOLE[40], is shown superimposed over the bisected map in translucent white with a navy outline. The HSP101 NBD2 pore loop densities are coloured by HSP101 protomer, and the cargo density is coloured pink. **e–j**, The transition from the asymmetric HSP101 spiral to the C7-pseudosymmetric PTEX150(668–823)–EXP2 heptamer is depicted using a series of cross-sections taken perpendicular to the central axis of the translocon, spanning the area of symmetry mismatch. The section of the translocon corresponding to each cross-sectional image is indicated with a bracket in **d**.





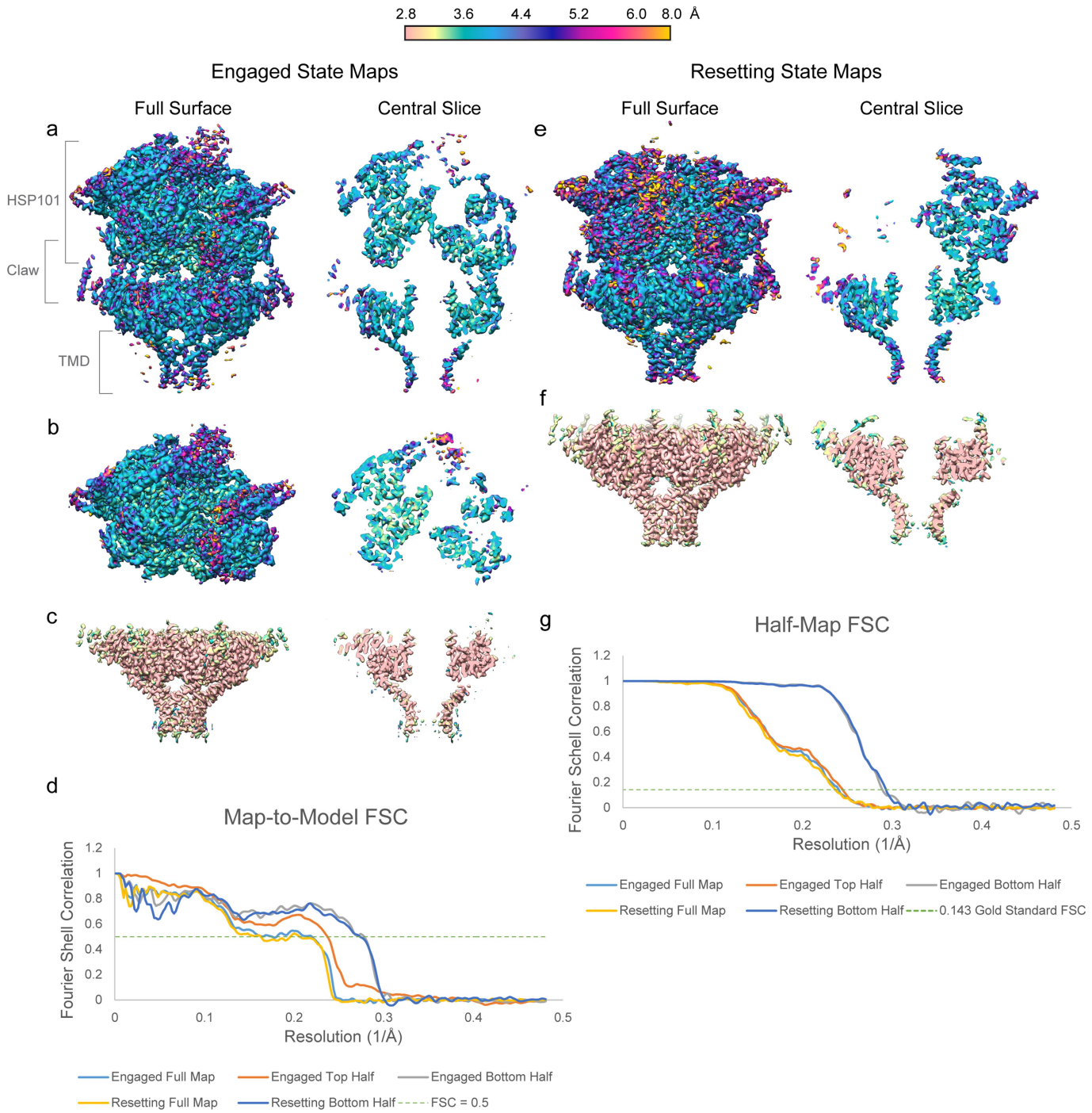




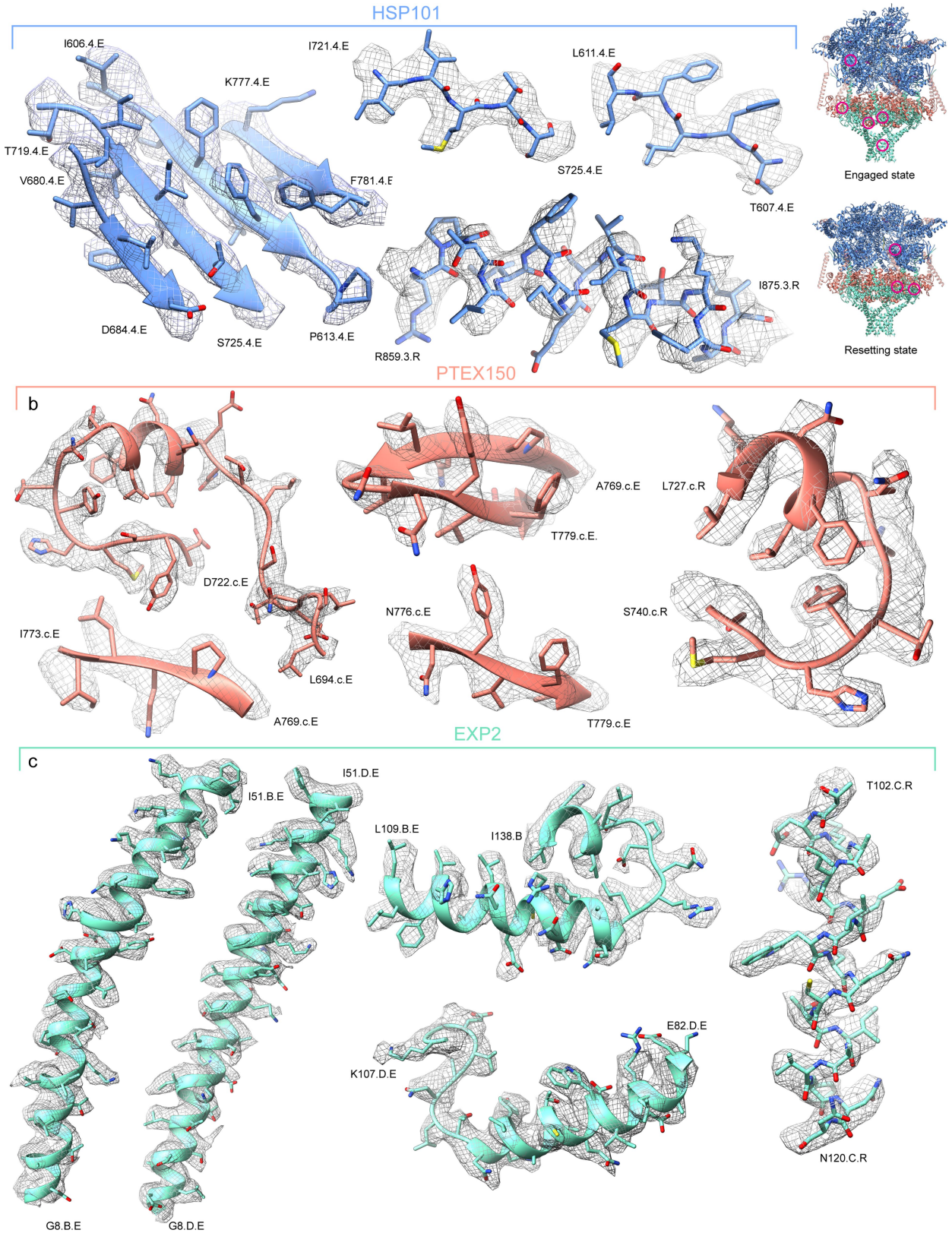


Extended Data Fig. 3 | See next page for caption.

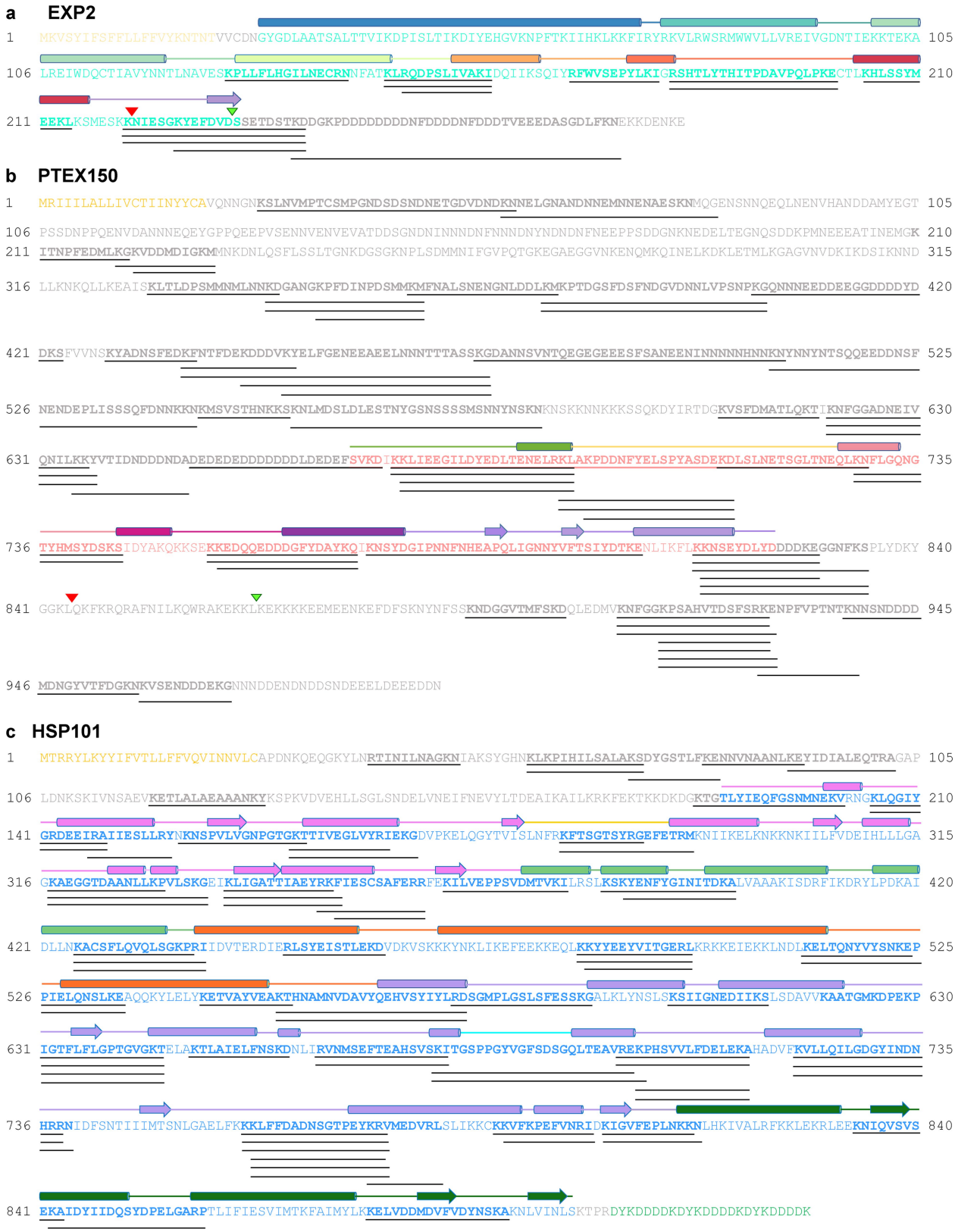
Extended Data Fig. 3 | Detailed comparisons of the engaged and resetting states. a, Side and top views of the EXP2 heptamer in the engaged state. Symmetric portions that remain constant between protomers are coloured in mint. Portions that vary between protomers are coloured and labelled by protomer. **b**, **c**, Superposition of the seven EXP2 protomers, labelled A–G, in the engaged (**b**) and resetting (**c**) states, coloured as in **a**. **d**, **e**, Top view of HSP101 NBD1 (**d**) and NBD2 (**e**) in the engaged and resetting states, shown in simplified surface representation. The hinge point at the interface between HSP101 protomers 3 and 4 is indicated. **f**, **g**, Ribbon diagrams of the resetting state (**f**) and engaged state (**g**) nucleotide binding pockets are shown for each protomer. ATPγS in each pocket is shown with corresponding cryo-EM density (mesh). The R859 arginine finger (sidechain shown in red-orange) is positioned approximately 3–5.5 Å from the phosphorus atom in the γ-phosphate of the ATPγS in the binding pocket of the neighbouring protomer in all protomers except R859 in protomer 3 in the resetting state (sidechain shown in gold), where the ATPγS bound in the protomer 4 NBD2 nucleotide pocket has shifted approximately 7.5 Å away from the protomer 3 R859 arginine finger. **h**, **i**, Enlarged side view of the atomic models of the HSP101 NBD2 pore loops and unfolded cargo polypeptide backbone in the engaged (**h**) and resetting (**i**) states, shown with corresponding cryo-EM densities. Tyrosine sidechain densities are clearly visibly intercalating with the cargo densities. The modelled PTEX cargo has a calculated r.m.s.d. of 1.09 Å and 1.25 Å to the published YME1 and HSP104 cargo models, respectively. Pore loops are labelled by NBD and protomer (for example, D2PL,P1 is NBD2 pore loop, protomer 1).



Extended Data Fig. 4 | Resolution assessments of the two PTEX states. **a**–**c**, **e**, **f**, Local resolution evaluations of the full PTEX map (**a**) and the focus-refined maps of the upper/hexameric (**b**) and lower/heptameric (**c**) halves of PTEX in the engaged state, the full PTEX map (**e**) and the focus-refined map of the lower/heptameric (**f**) half of PTEX in the resetting state, calculated by Resmap[56] and coloured according to resolution. Maps are displayed at higher thresholds where the detergent belt is not visible for clarity, to avoid obscuring details of the transmembrane helices. **d**, Global resolution assessment of the engaged and resetting state maps as measured using the 'Gold-standard' Fourier shell correlation (FSC) curves generated by RELION[44,45] by comparison of two independently refined 'half-maps'. **g**, Map-to-model FSC curves demonstrating the degree of correlation between the refined PTEX models and the experimental cryo-EM maps for the engaged and resetting states.

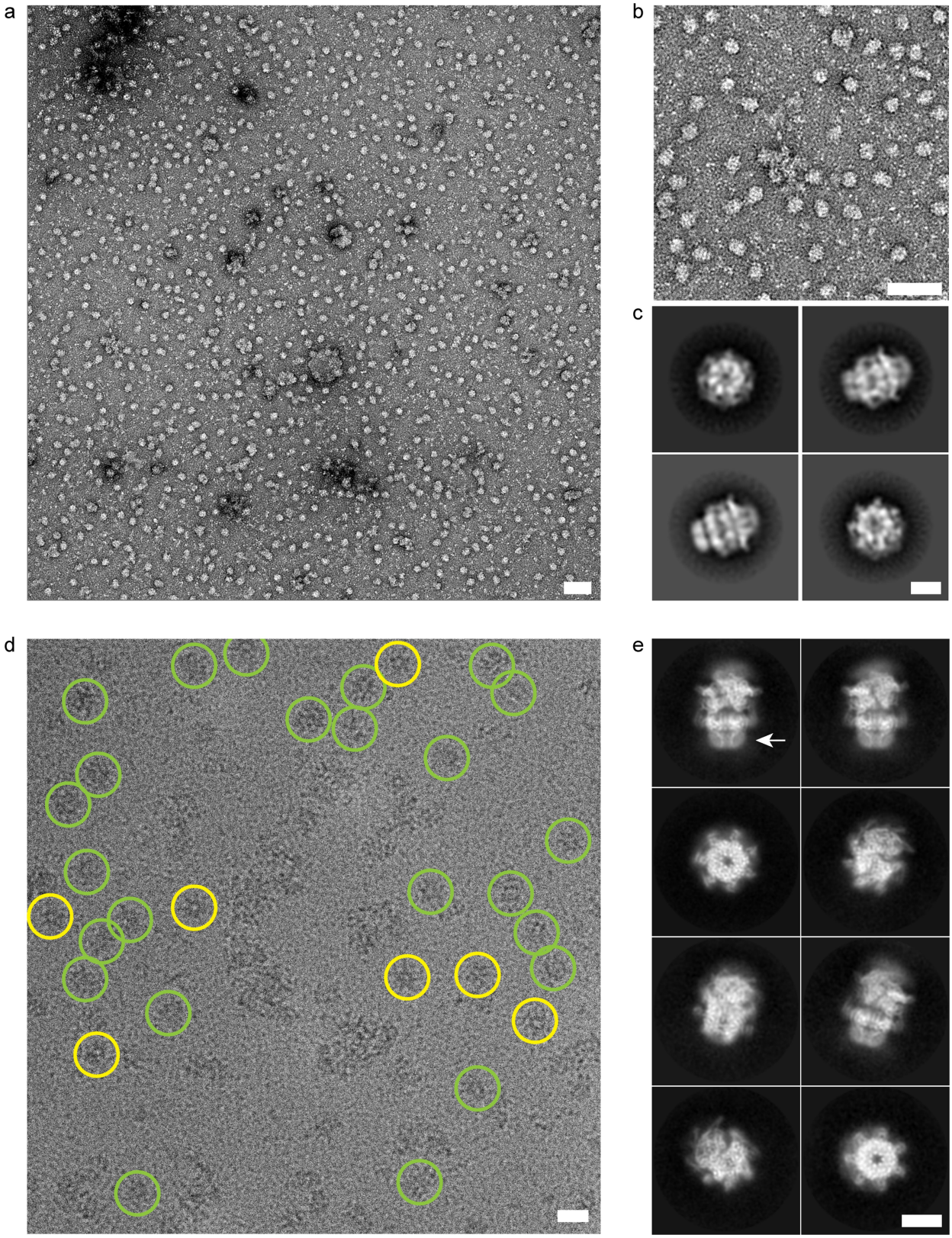


Extended Data Fig. 5 | Representative regions of cryo-EM density and atomic models. Additional cryo-EM densities (mesh) superposed with our atomic models for HSP101 (**a**), PTEX150(668–823) (**b**) and EXP2 (**c**). Displayed regions correspond to areas circled in magenta on guide figures (inset, upper right), and are coloured as in the guide figure: HSP101 (cornflower blue), PTEX150(668-823) (salmon), EXP2 (mint). Terminal residues for each segment are labelled with the amino acid, residue number, protomer, and state.

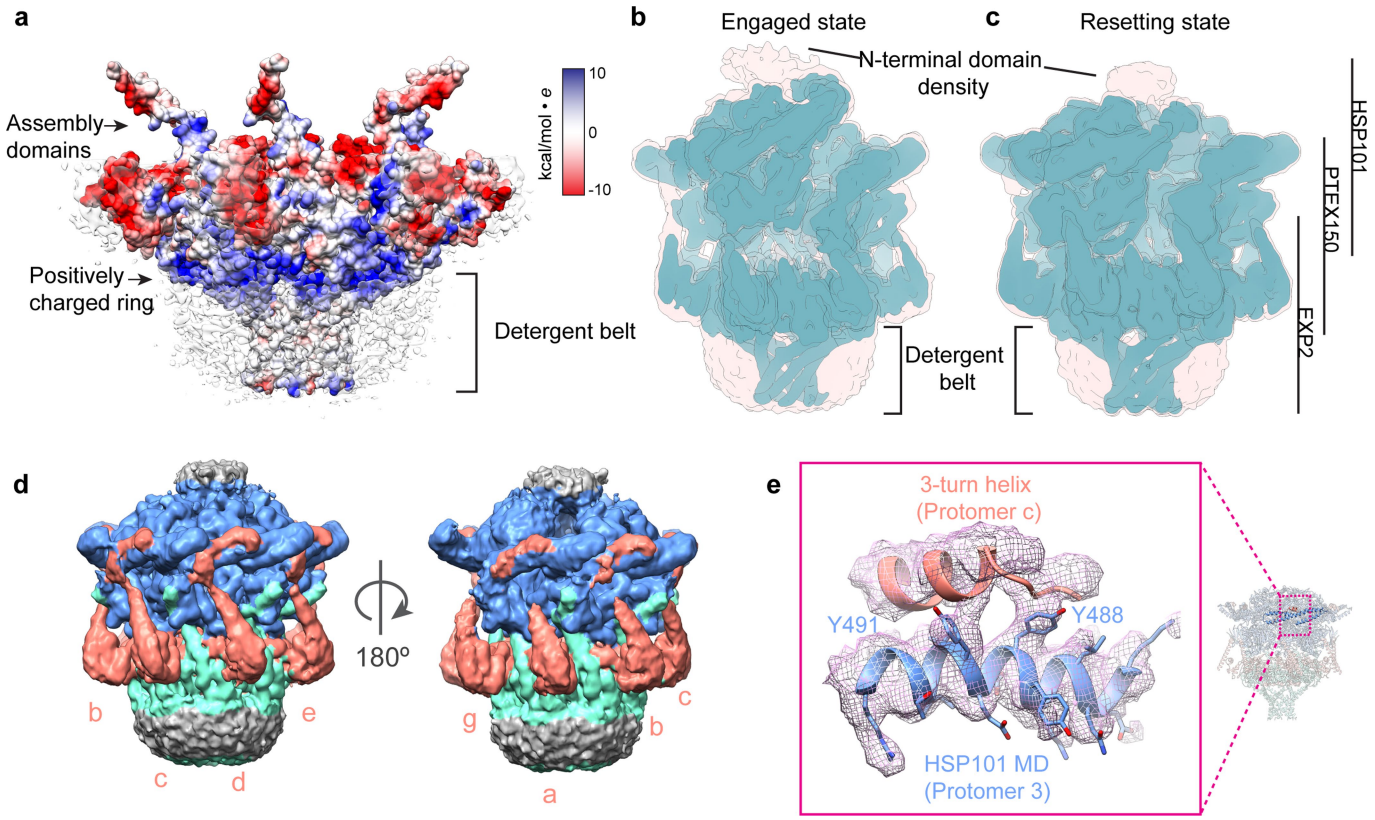


Extended Data Fig. 6 | See next page for caption.

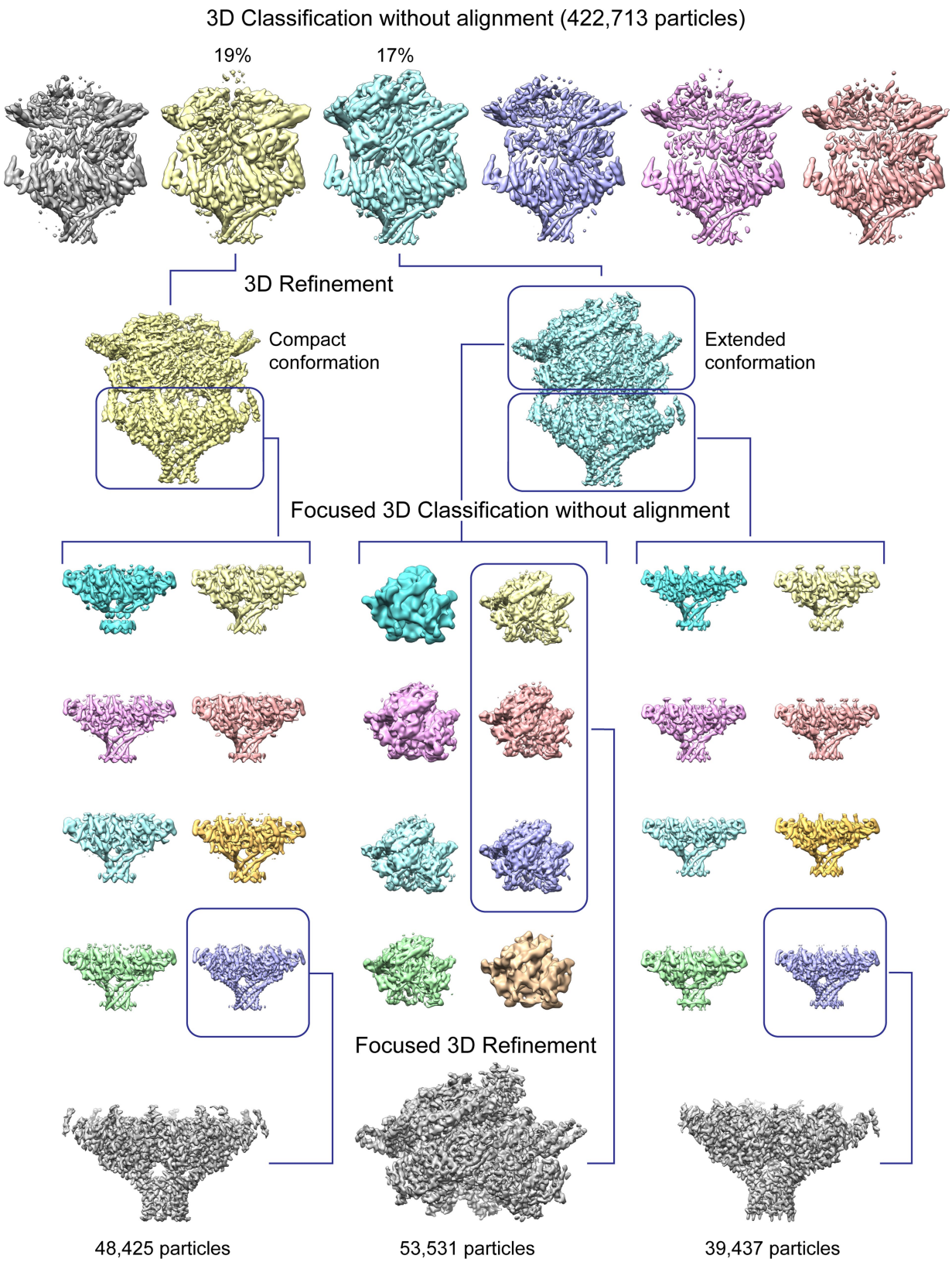
Extended Data Fig. 6 | Experimentally determined secondary structure elements and detected mass-spectrometry fragments mapped to the primary sequences of the three PTEX proteins. For EXP2 (**a**), PTEX150 (**b**) and HSP101 (**c**), secondary structure elements are shown as tubes (helices), lines (loops), and arrows (strands) above the corresponding sequence and are coloured as in Figs. 2a, 3a, 4a. In the sequences shown below, residues resolved in our structures are coloured according to protein colours in Fig. 1c–f: EXP2 (mint), PTEX150 (salmon) and HSP101 (cornflower). Signal peptide residues are coloured gold. All residues in the mature proteins that are not resolved in our structures are shown in grey. The 3×Flag residues at the C terminus of HSP101 are coloured green. Peptides detected in tryptic digest liquid chromatography–tandem mass spectrometry (LC–MS/MS) analysis of the purified PTEX sample are shown as black lines below the corresponding sequences. Arrowheads above the EXP2 sequence indicate truncations sites described in this work and previously[20] immediately before (Δ222–287, red arrowhead) and after (Δ234–287, green arrowhead) the assembly strand. Arrowheads above the PTEX150 sequence indicate previously described truncation sites[37] (Δ847–993, red arrowhead; Δ869–993, green arrowhead).



Extended Data Fig. 7 | Electron microscopy of the PTEX core complex. **a**–**c**, Representative negative stain micrograph (**a**), enlarged portion of micrograph (**b**) and two-dimensional class averages (**c**) of the PTEX core complex in multiple orientations. **d**, **e**, Representative cryo-EM micrograph (**d**) and two-dimensional class averages (**e**) of the PTEX core complex in multiple orientations. Individual particles in **d** are circled in yellow (top views) and green (side and oblique views) (for source data, see Supplementary Fig. 4). Arrow in upper left panel of **e** indicates the detergent belt, which is visible as a less-dense (dimmer) halo surrounding the denser (brighter) densities of the α-helices visible in the transmembrane domain in side views. Scale bars: 700 Å (**a**), 700 Å (**b**), 100 Å (**c**), 200 Å (**d**) and 100 Å (**e**).



Extended Data Fig. 8 | Detergent belt, amino-terminal domain, and claw densities visible in maps at lower thresholds. a, The engaged state PTEX150–EXP2 heptamer, displayed in surface representation and coloured by electrostatic potential. The bottom half of the full engaged state density map is superimposed, showing the location of the detergent belt in relation to the EXP2 transmembrane domain. A ring of positively charged residues is clearly visible directly above where the PVM surface would normally lie. **b**, **c**, Engaged state (**b**) and resetting state (**c**) maps were low-pass filtered to 6 Å to improve clarity of low-resolution details, and are shown overlaid, at two different thresholds to improve visibility of the detergent belt and the poorly resolved N-terminal domains of HSP101 (teal, higher threshold; peach, lower threshold). **d**, Resetting state map of PTEX displayed at a lower threshold to show the strong claw-shaped densities extending from the PTEX150(668–823) shaft up to the HSP101 M domain. **e**, Enlarged view of the interaction between HSP101 Y488 and Y491 and the three-turn helix, shown with corresponding cryo-EM density (mesh).



Extended Data Fig. 9 | Data processing workflow. Illustration of workflow for 3D classification, focused classification and refinement. Maps are displayed at higher thresholds where the detergent belt is not visible for clarity, to avoid obscuring details of the transmembrane helices.

Extended Data Table 1 | Cryo-EM data collection, refinement and validation statistics

	PTEX *Engaged* Full (EMDB-8951) (PDB 6E10)	PTEX *Engaged* Top (EMDB-8951) (PDB 6E10)	PTEX *Engaged* Bottom (EMDB-8951) (PDB 6E10)	PTEX *Resetting* Full (EMDB-8952) (PDB 6E11)	PTEX *Resetting* Top (EMDB-8952) (PDB 6E11)	PTEX *Resetting* Bottom (EMDB-8952) (PDB 6E11)
Data collection and processing						
Magnification	×105,000	×105,000	×105,000	×105,000	n/a	×105,000
Voltage (kV)	300	300	300	300	n/a	300
Electron exposure (e–/Å²)	60	60	60	60	n/a	60
Defocus range (μm)	-1.5 to -4.0	-1.5 to -4.0	-1.5 to -4.0	-1.5 to -4.0	n/a	-1.5 to -4.0
Pixel size (Å)	1.04	1.04	1.04	1.04	n/a	1.04
Symmetry imposed	C1	C1	C7	C1	n/a	C7
Initial particle images (no.)	1,508,462	1,508,462	1,508,462	1,508,462	n/a	1,508,462
Final particle images (no.)	72,866	53,531	39,437	78,499	n/a	48,425
Map resolution (Å) FSC threshold	4.09 0.143	4.16 0.143	3.5 0.143	4.23 0.143	n/a	3.4 0.143
Map resolution range (Å)	3.2-7.5	3.0-7.0	2.8-3.6	3.2-7.5	n/a	2.8-3.4
Refinement						
Initial model used (PDB code)	n/a	n/a	n/a	n/a	n/a	n/a
Model resolution (Å) FSC threshold	4.58 0.5	4.23 0.5	3.59 0.5	4.84 0.5	n/a	3.67 0.5
Model resolution range (Å)	4.58	4.23	3.59	4.84	n/a	3.67
Map sharpening *B* factor (Å²)	-180	-180	-170	-180	n/a	-160
Model composition Non-hydrogen atoms Protein residues Ligands	 57,352 6,838 12	 57,352 6,838 12	 57,352 6,838 12	 57,401 6,826 12	 57,401 6,826 12	 57,401 6,826 12
B factors (Å²) Protein Ligand	n/a	n/a	n/a	n/a	n/a	n/a
R.m.s. deviations Bond lengths (Å) Bond angles (°)	 0.008 1.311	 0.009 1.262	 0.008 0.925	 0.008 1.300	 0.007 1.332	 0.006 0.895
Validation MolProbity score Clashscore Poor rotamers (%)	 1.95 9.74 1.04	 1.81 7.35 0.13	 1.53 4.01 0.45	 2.03 10.44 0.47	 1.96 7.46 0.05	 1.64 5.48 0.33
Ramachandran plot Favored (%) Allowed (%) Disallowed (%)	 93.96 5.53 0.51	 94.01 5.76 0.23	 94.99 4.69 0.33	 91.91 7.65 0.44	 89.98 9.58 0.44	 94.98 5.08 0.00

LETTER

https://doi.org/10.1038/s41586-018-0468-5

A complex dynamo inferred from the hemispheric dichotomy of Jupiter's magnetic field

Kimberly M. Moore[1], Rakesh K. Yadav[1], Laura Kulowski[1], Hao Cao[1], Jeremy Bloxham[1]*, John E. P. Connerney[2,3], Stavros Kotsiaros[2,4], John L. Jørgensen[5], José M. G. Merayo[5], David J. Stevenson[6], Scott J. Bolton[7] & Steven M. Levin[8]

The Juno spacecraft, which is in a polar orbit around Jupiter, is providing direct measurements of the planet's magnetic field close to its surface[1]. A recent analysis of observations of Jupiter's magnetic field from eight (of the first nine) Juno orbits has provided a spherical-harmonic reference model (JRM09)[2] of Jupiter's magnetic field outside the planet. This model is of particular interest for understanding processes in Jupiter's magnetosphere, but to study the field within the planet and thus the dynamo mechanism that is responsible for generating Jupiter's main magnetic field, alternative models are preferred. Here we report maps of the magnetic field at a range of depths within Jupiter. We find that Jupiter's magnetic field is different from all other known planetary magnetic fields. Within Jupiter, most of the flux emerges from the dynamo region in a narrow band in the northern hemisphere, some of which returns through an intense, isolated flux patch near the equator. Elsewhere, the field is much weaker. The non-dipolar part of the field is confined almost entirely to the northern hemisphere, so there the field is strongly non-dipolar and in the southern hemisphere it is predominantly dipolar. We suggest that Jupiter's dynamo, unlike Earth's, does not operate in a thick, homogeneous shell, and we propose that this unexpected field morphology arises from radial variations, possibly including layering, in density or electrical conductivity, or both.

Unlike Earth, for which the top of the dynamo region is well defined by the core–mantle boundary—that is, the boundary between the electrically conducting liquid-iron outer core (in which dynamo action occurs) and the overlying, poorly conducting rocky mantle—for Jupiter the corresponding region is less clearly defined. Even though self-sustaining dynamo action is most probably confined to depths below the metallic-hydrogen transition, the field may be affected by flow in the overlying molecular-hydrogen region[3–5], which may have substantial electrical conductivity, especially close to the depth of the metallic-hydrogen transition[6,7]. Accordingly, we map the field at four equally spaced radii from the surface of Jupiter (corresponding to $r = R_J = 71{,}492$ km, where R_J is Jupiter's radius), at which the electrical conductivity is vanishingly small, to $r = 0.85R_J$, the likely depth of the metallic-hydrogen transition.

To do so requires mapping the field below the orbit of the spacecraft, and so we must address the instability due to downward continuation. We do so by regularizing the solution using a quadratic norm based on the horizontal Laplacian of the radial magnetic field, thereby finding the smoothest possible map of the field for a given fit to the observations[8]. We select Juno magnetometer observations[1] from eight orbits in the radial distance range from $r = 1.06R_J$ (perijove) to $r = 2.2R_J$ (roughly corresponding to Juno's highest latitude), take 30-s averages of the data (corresponding to one rotation of the spacecraft) and weight the data according to an estimate of their measurement uncertainty. Our resulting dataset consists of 1,991 observations of each of the three components of the magnetic field.

In Fig. 1 we show maps of the radial component of the magnetic field at a range of depths using our regularized inversion from the surface to $r = 0.85R_J$ and compare with JRM09[2]. At all depths, positive radial flux in the northern hemisphere is confined to a band (the northern-hemisphere flux band), which becomes narrower with depth. Some of the flux from this band then re-enters through an intense spot on the equator[9] (the Great Blue Spot), at a longitude of around 90° west (in System III coordinates). The morphology of the magnetic field lines is shown in Fig. 2 (an animated version of Fig. 2 is available at https://doi.org/10.6084/m9.figshare.6828953). Elsewhere, and corresponding to a large proportion of the surface, the radial flux is much weaker.

The narrowing of the northern-hemisphere flux band with depth, and more generally the concentration of flux into increasingly localized regions with depth rather than, for example, the emergence of more small-scale spots, is surprising given our intuition acquired from mapping Earth's magnetic field at depth. It suggests that Jupiter's magnetic field at depth may be morphologically simpler than expected. This field morphology and its contrast to Earth's field is particularly apparent in Fig. 3, in which we show the non-dipolar part of the field (at $r = 0.90R_J$) and, for comparison, Earth's non-dipole field (at Earth's core–mantle boundary). Jupiter's non-dipole field is almost entirely confined to the northern hemisphere, where the non-dipole field peaks at 3 mT, a value almost three times stronger than the peak dipolar field. Jupiter's field is dipolar in the southern hemisphere and largely non-dipolar in the northern hemisphere, unlike Earth's field.

The strong concentration of magnetic flux in the northern-hemisphere flux band and in the Great Blue Spot implies the existence of large horizontal magnetic field gradients at the borders of these features, which would suggest that strong secular (temporal) variation of the magnetic field is likely. For example, around the Great Blue Spot the gradient in the radial field is approximately 3 mT/(10^6 m); with an assumed flow speed of the order of 10^{-4} m s^{-1} (the lower end of estimates of flow speed[10,11]), we might therefore expect secular variation of the order of 10^4 nT yr^{-1}. Although high, this estimate is not necessarily inconsistent with earlier inferences of much weaker time dependency[12] because secular variation at such small spatial scales would be strongly attenuated at the altitude of the previous observations. In addition, this estimate will be reduced if the flow is preferentially orthogonal to the field gradient, although for the Great Blue Spot that is unlikely on geometrical grounds to be the case. Therefore, we believe that the Great Blue Spot offers a very promising opportunity for forthcoming Juno orbits to detect secular variation.

Numerical dynamo models in simple homogeneous shells typically produce fields that are either strongly dipolar or dominated by multipolar fields[10,13]. Jupiter's field is neither, being predominantly dipolar in one hemisphere and non-dipolar in the other, suggesting that the field is not generated in a simple homogeneous region. Here we consider several possible explanations. First, we consider the possibility, although unlikely, that we have observed the field in a rare transitional

[1]Department of Earth and Planetary Sciences, Harvard University, Cambridge, MA, USA. [2]NASA/GSFC, Greenbelt, MD, USA. [3]Space Research Corporation, Annapolis, MD, USA. [4]University of Maryland, College Park, MD, USA. [5]National Space Institute, Technical University of Denmark, Kongens Lyngby, Denmark. [6]Division of Geological and Planetary Sciences, California Institute of Technology, Pasadena, CA, USA. [7]Southwest Research Institute, San Antonio, TX, USA. [8]Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA, USA.
*e-mail: jeremy_bloxham@harvard.edu

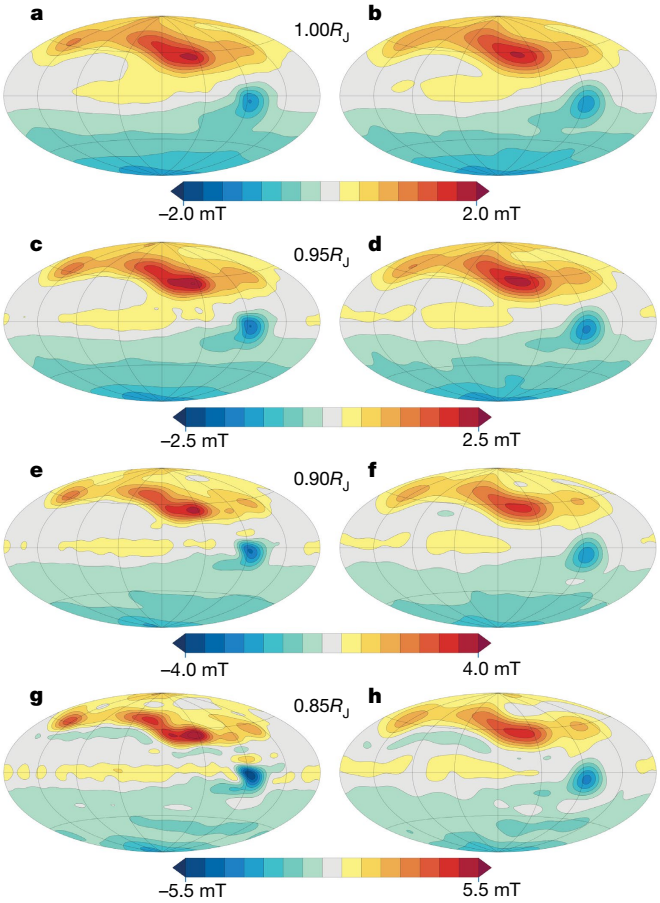


Fig. 1 | The radial component of Jupiter's magnetic field. The plots are shown on a Hammer equal-area projection with the central meridian at a longitude of 180° west (System III coordinates). The colour scale depicts the strength of the radial component of the magnetic field, with yellow–red shades depicting field in the positive radial direction (outwards) and green–blue shades depicting field in the negative radial direction (inwards). **a**, **b**, A regularized solution (**a**) and the JRM09 solution (**b**) at $r = 1.00R_J$; **c**, **d**, the same at $r = 0.95R_J$; **e**, **f**, the same at $r = 0.90R_J$; **f**, **g**, the same at $r = 0.85R_J$. Although the regularized solution and the JRM09 solution have a similar pattern at each depth, the regularized solution reveals more intense and concentrated field structure. Overall, the same basic field morphology is apparent across the range of depths and the two models.

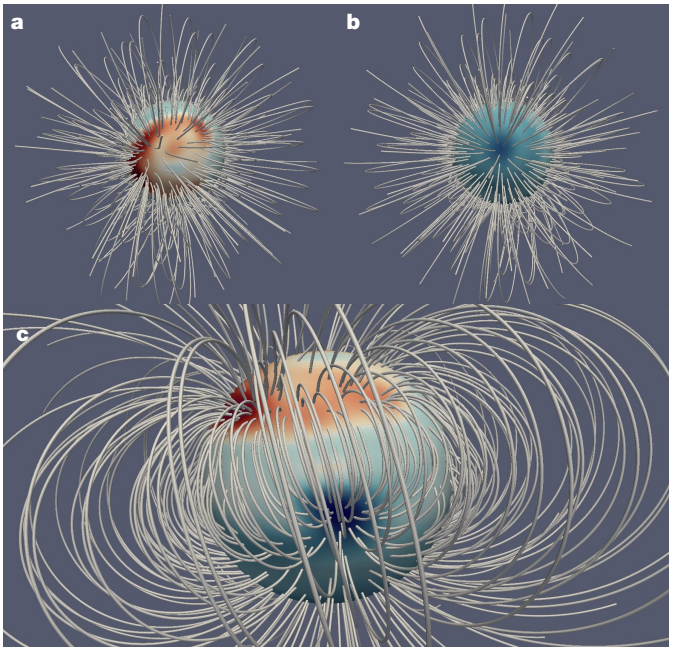


Fig. 2 | Magnetic field lines. a, North polar view; **b**, south polar view; **c**, equatorial view. The non-dipolar nature of the magnetic field in the northern hemisphere and the dipolar nature in the southern hemisphere is apparent. The equatorial view is centred near the Great Blue Spot and shows the linkage of magnetic field lines that enter through the Great Blue Spot. The contoured surface on which the field lines shown start and end is at $r = 0.85R_J$, where the density of field lines is proportional to the radial magnetic field strength and is depicted by the colour scale (red outward flux, blue inward flux). An animated version of this figure is available at https://doi.org/10.6084/m9.figshare.6828953.

state, such as a magnetic field reversal or a transition between different dynamo states[14,15]. However, such a situation cannot necessarily be reconciled with the co-existence of strong dipole and non-dipole fields. Instead, we next consider whether Jupiter's internal structure could account for the observations.

Starting near the top of Jupiter's dynamo region, there is the possibility of a stably stratified layer due to precipitation of helium[16]. Such a layer might axisymmetrize the field[17], but could also destabilize the field[18]. However, this scenario also seems unlikely to be able to account for the observed hemispheric difference in the field morphology. There is also the effect of the steep gradient in electrical conductivity immediately above the metallic-hydrogen transition[7]. A recent numerically simulated dynamo including this effect shows irregular behaviour[19], with one snapshot appearing similar to the Juno-determined field. This is a possibility that requires further investigation. Finally, another recent study[20] has examined flow and the generation of magnetic fields in Jupiter for three scenarios that involve near-surface layering, although none of the scenarios produces magnetic fields similar to that observed by Juno.

At depth, other processes may be important. In particular, the mixture of rock and ice that probably constitutes (or constituted) Jupiter's core will be soluble in hydrogen at the temperature and pressure expected there[21–25]. This may lead to gradual core dissolution, and may have been crucial in Jupiter's thermal history[26,27]. Dissolution of rock and ice in metallic hydrogen will increase the density of the hydrogen region. Recent Juno observations of Jupiter's gravity field are consistent with the existence of a partially or fully dissolved core inside Jupiter, with rock and ice non-uniformly mixed in the hydrogen out to approximately half the radius of the planet[28]; the region further out may be homogeneous, except for helium rain.

If, as theory and observations suggest, the metallic-hydrogen region is layered (the upper layer solute-free and the lower layer containing dissolved rock and ice), the implications for the dynamo will depend on the convective instability of these layers. The upper layer is most probably convectively unstable, given the very large heat flux observed at Jupiter. The properties of the lower layer are far less clear. If the lower layer is stable, then dynamo action will be confined to the upper layer and will therefore operate in a shell with a radius ratio (inner to outer radii) of approximately 0.5. A similar geometry has been investigated previously as a possible explanation for the magnetic fields of Uranus and Neptune[29], albeit with a numerical dynamo model much less sophisticated than what is now feasible. The magnetic field map obtained from this simulation with a radius ratio of 0.5 (see figure 16, model 5 in ref. 29) bears similarity to the map of Jupiter's field shown here, but with an axial dipole that is much less dominant. In addition, structure may arise from double diffusive convection[26].

Alternatively, if the lower layer is convectively unstable, then it could be convecting separately from the layer above owing to the possible presence of a density jump at the boundary between the layers[28]. Convection in Jupiter's metallic-hydrogen region can be driven by relative density variations ($\Delta\rho/\rho$) of the order of 10^{-6}, so even a small density jump could be impervious to convection. In this scenario, dynamo action may occur separately in the thick lower shell

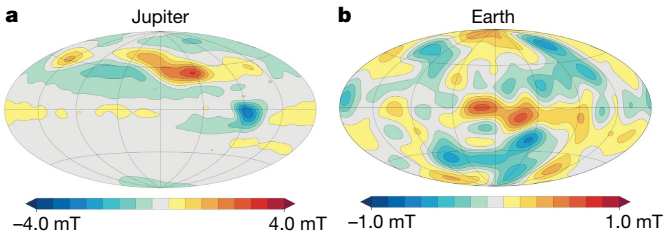


Fig. 3 | Non-dipole radial field. a, The non-dipolar part of Jupiter's radial magnetic field at $r = 0.90R_J$. **b**, For comparison, the non-dipolar part of Earth's radial magnetic field at the core–mantle boundary ($r = 0.55R_E = 3{,}485$ km, where R_E is Earth's radius). Almost all of Jupiter's non-dipole radial field is concentrated in the northern hemisphere, whereas Earth's field is evenly distributed throughout.

(radius ratio of less than 0.2) and in the thin upper shell (radius ratio of approximately 0.5), with the resultant field sharing properties of both a thick-shell dynamo (strong axial dipole) and a relatively thin-shell dynamo (hemispheric asymmetry).

The presence or absence of reduced magnetic flux at high latitude may provide a means of distinguishing between these alternatives. If the lower layer is stably stratified, then convection in the outer layer within the tangent cylinder (the axial cylinder tangential to the interface between the two layers) may differ from that outside the tangent cylinder. If the lower layer is convectively unstable, then such an effect seems less likely to occur. To resolve this additional Juno orbits are required. Juno's orbit, with perijove precessing northward by approximately 1° per orbit, is evolving in such a way that mid- and high-latitude structure will be better resolved towards the second half of the planned 34-orbit baseline mission[30].

Data availability

The Juno magnetometer data used in this study will be made available through the NASA Planetary Data System (https://pds.nasa.gov) in accordance with NASA policy. An animated version of Fig. 2 is available at https://doi.org/10.6084/m9.figshare.6828953.

Received: 9 April 2018; Accepted: 26 July 2018;
Published online 5 September 2018.

1. Connerney, J. E. P. et al. The Juno magnetic field investigation. *Space Sci. Rev.* **213**, 39–138 (2017).
2. Connerney, J. E. P. et al. A new model of Jupiter's magnetic field from Juno's first nine orbits. *Geophys. Res. Lett.* **45**, 2590–2596 (2018).
3. Liu, J., Goldreich, P. M. & Stevenson, D. J. Constraints on deep-seated zonal winds inside Jupiter and Saturn. *Icarus* **196**, 653–664 (2008).
4. Gastine, T., Wicht, J., Duarte, L., Heimpel, M. & Becker, A. Explaining Jupiter's magnetic field and equatorial jet dynamics. *Geophys. Res. Lett.* **41**, 5410–5419 (2014).
5. Cao, H. & Stevenson, D. J. Zonal flow magnetic field interaction in the semi-conducting region of giant planets. *Icarus* **296**, 59–72 (2017).
6. Nellis, W. J., Weir, S. T. & Mitchell, A. C. Metallization and electrical conductivity of hydrogen in Jupiter. *Science* **273**, 936–938 (1996).
7. French, M. et al. Ab initio simulations for material properties along the Jupiter adiabat. *Astrophys. J. Suppl. Ser.* **202**, 5 (2012).
8. Shure, L., Parker, R. L. & Backus, G. E. Harmonic splines for geomagnetic modelling. *Phys. Earth Planet. Inter.* **28**, 215–229 (1982).
9. Moore, K. M., Bloxham, J., Connerney, J. E. P., Jørgensen, J. L. & Merayo, J. M. G. The analysis of initial Juno magnetometer data using a sparse magnetic field representation. *Geophys. Res. Lett.* **44**, 4687–4693 (2017).
10. Christensen, U. R. & Aubert, J. Scaling properties of convection-driven dynamos in rotating spherical shells and application to planetary magnetic fields. *Geophys. J. Int.* **166**, 97–114 (2006).
11. Jones, C. A. A dynamo model of Jupiter's magnetic field. *Icarus* **241**, 148–159 (2014).
12. Ridley, V. A. & Holme, R. Modeling the Jovian magnetic field and its secular variation using all available magnetic field observations. *J. Geophys. Res. Planets* **121**, 309–337 (2016).
13. Jones, C. A. Planetary magnetic fields and fluid dynamos. *Annu. Rev. Fluid Mech.* **43**, 583–614 (2011).
14. Duarte, L. D. V., Wicht, J. & Gastine, T. Physical conditions for Jupiter-like dynamo models. *Icarus* **299**, 206–221 (2018).
15. Grote, E. & Busse, F. H. Hemispherical dynamos generated by convection in rotating spherical shells. *Phys. Rev. E* **62**, 4457–4460 (2000).
16. Salpeter, E. E. On convection and gravitational layering in Jupiter and in stars of low mass. *Astrophys. J.* **181**, L83–L86 (1973).
17. Stevenson, D. J. Reducing the non-axisymmetry of a planetary dynamo and an application to Saturn. *Geophys. Astrophys. Fluid Dyn.* **21**, 113–127 (1982).
18. Stanley, S. & Mohammadi, A. Effects of an outer thin stably stratified layer on planetary dynamos. *Phys. Earth Planet. Inter.* **168**, 179–190 (2008).
19. Dietrich, W. & Jones, C. A. Anelastic spherical dynamos with radially variable electrical conductivity. *Icarus* **305**, 15–32 (2018).
20. Glatzmaier, G. A. Computer simulations of Jupiter's deep internal dynamics help interpret what Juno sees. *Proc. Natl Acad. Sci. USA* **115**, 6896–6904 (2018).
21. Stevenson, D. J. Cosmochemistry and structure of the giant planets and their satellites. *Icarus* **62**, 4–15 (1985).
22. Wilson, H. F. & Militzer, B. Rocky core solubility in Jupiter and giant exoplanets. *Phys. Rev. Lett.* **108**, 111101 (2012).
23. Wilson, H. F. & Militzer, B. Solubility of water ice in metallic hydrogen: consequences for core erosion in gas giant planets. *Astrophys. J.* **745**, 54 (2012).
24. Wahl, S. M., Wilson, H. F. & Militzer, B. Solubility of iron in metallic hydrogen and stability of dense cores in giant planets. *Astrophys. J.* **773**, 95 (2013).
25. González-Cataldo, F., Wilson, H. F. & Militzer, B. Ab initio free energy calculations of the solubility of silica in metallic hydrogen and application to giant planet cores. *Astrophys. J.* **787**, 79 (2014).
26. Helled, R. & Stevenson, D. The fuzziness of giant planets' cores. *Astrophys. J. Lett.* **840**, L4 (2017).
27. Vazan, A., Helled, R. & Guillot, T. Jupiter's evolution with primordial composition gradients. *Astron. Astrophys.* **610**, L14 (2018).
28. Wahl, S. M. et al. Comparing Jupiter interior structure models to Juno gravity measurements and the role of a dilute core. *Geophys. Res. Lett.* **44**, 4649–4659 (2017).
29. Stanley, S. & Bloxham, J. Numerical dynamo models of Uranus' and Neptune's magnetic fields. *Icarus* **184**, 556–572 (2006).
30. Bolton, S. J. et al. The Juno mission. *Space Sci. Rev.* **213**, 5–37 (2017).

Acknowledgements All authors acknowledge support from the Juno project. K.M.M. is supported by the US Department of Defense (DoD) through the National Defense Science and Engineering Graduate Fellowship (NDSEG) programme and L.K. through a US National Science Foundation Graduate Fellowship.

Reviewer information *Nature* thanks C. Jones and the other anonymous reviewer(s) for their contribution to the peer review of this work.

Author contributions K.M.M. and J.B. wrote the manuscript and performed the data analysis. K.M.M., J.B., J.E.P.C., S.K., J.L.J. and J.M.G.M. contributed to discussions of the data analysis, and K.M.M., R.K.Y., L.K., H.C., J.B. and D.J.S. contributed to discussions of the dynamo implications. All authors contributed to editing and revising the manuscript. J.E.P.C. is principal investigator of the Juno magnetometer investigation, S.J.B. is principal investigator of the mission and S.M.L. is project scientist of the mission.

Competing interests The authors declare no competing interests.

Additional information
Reprints and permissions information is available at http://www.nature.com/reprints.
Correspondence and requests for materials should be addressed to J.B.
Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

LETTER

https://doi.org/10.1038/s41586-018-0450-2

Synthetic three-dimensional atomic structures assembled atom by atom

Daniel Barredo[1,2]*, Vincent Lienhard[1,2], Sylvain de Léséleuc[1,2], Thierry Lahaye[1] & Antoine Browaeys[1]

A great challenge in current quantum science and technology research is to realize artificial systems of a large number of individually controlled quantum bits for applications in quantum computing and quantum simulation. Many experimental platforms are being explored, including solid-state systems, such as superconducting circuits[1] or quantum dots[2], and atomic, molecular and optical systems, such as photons, trapped ions or neutral atoms[3–7]. The latter offer inherently identical qubits that are well decoupled from the environment and could provide synthetic structures scalable to hundreds of qubits or more[8]. Quantum-gas microscopes[9] allow the realization of two-dimensional regular lattices of hundreds of atoms, and large, fully loaded arrays of about 50 microtraps (or 'optical tweezers') with individual control are already available in one[10] and two[11] dimensions. Ultimately, however, accessing the third dimension while keeping single-atom control will be required, both for scaling to large numbers and for extending the range of models amenable to quantum simulation. Here we report the assembly of defect-free, arbitrarily shaped three-dimensional arrays, containing up to 72 single atoms. We use holographic methods and fast, programmable moving tweezers to arrange—atom by atom and plane by plane—initially disordered arrays into target structures of almost any geometry. These results present the prospect of quantum simulation with tens of qubits arbitrarily arranged in space and show that realizing systems of hundreds of individually controlled qubits is within reach using current technology.

Three-dimensional atomic arrays at half filling have been obtained using optical lattices with large spacings[12], which facilitate single-site addressability and atom manipulation[13]. As an alternative approach, here we use programmable holographic optical tweezers to create three-dimensional (3D) arrays of traps. Holographic methods offer the advantage of higher tunability of the lattice geometry because the design of optical potential landscapes is reconfigurable and only limited by diffraction[14–16]. In our experiment[14], arbitrarily designed arrays of up to about 120 traps are generated by imprinting a phase pattern on a dipole trap beam at 850 nm with a spatial light modulator (Fig. 1a). This phase mask is calculated using the 3D Gerchberg–Saxton algorithm, simplified for the case of point traps[17]. The beam is then focused with a high-numerical-aperture (0.5) aspheric lens under vacuum, creating individual optical tweezers with a measured $1/e^2$ radius of about 1.1 μm and a Rayleigh length of approximately 5 μm. After recollimation with a second aspheric lens, the intensity of the trapping light is measured using a standard charge-coupled device (CCD) camera. An electrically tunable lens (ETL1) in the imaging path allows us to acquire series of stack images along the optical axis z, from which we reconstruct the full 3D intensity distribution. The imaging system covers a z-direction scan range of 200 μm.

Figure 1b–d shows some examples of patterns suitable for experiments with single atoms. The images are reconstructed using a maximum-intensity projection method[18] from 200 z images obtained with the diagnostics CCD camera. With about 3.5 mW of power per trap we reach depths of $U_0/k_B \approx 1$ mK, where k_B is the Boltzmann constant, and radial (longitudinal) trapping frequencies of around 100 kHz (20 kHz). We produce highly uniform microtrap potentials (with peak intensities differing by less than 5% root mean square) via a closed-loop

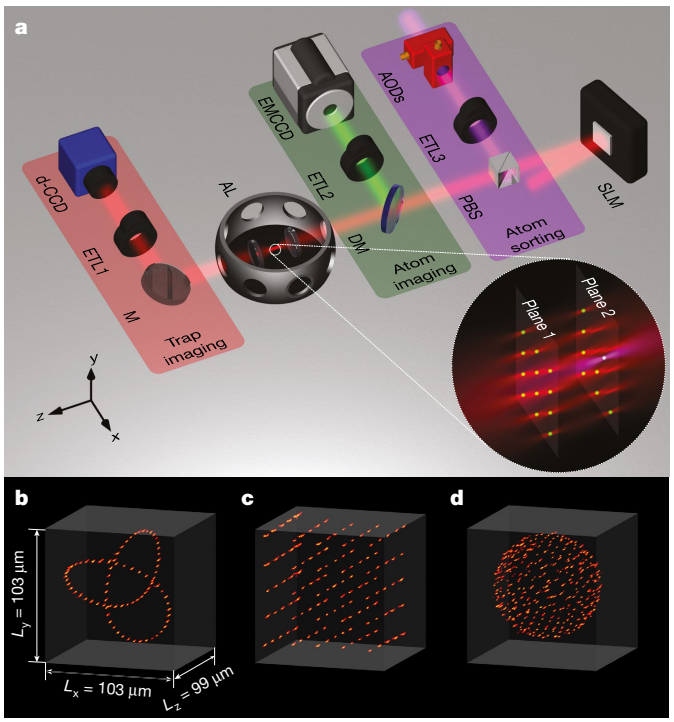


Fig. 1 | Experimental setup and trap images. a, We combine a spatial light modulator (SLM) and a high-numerical-aperture aspheric lens (AL) under vacuum to generate arbitrary 3D arrays of traps. The intensity distribution in the focal plane is measured with the aid of a second aspheric lens, a mirror (M) and a diagnostics CCD camera (d-CCD). The fluorescence of the atoms in the traps at 780 nm is separated from the dipole trap beam with a dichroic mirror (DM) and detected using an electron-multiplying CCD camera (EMCCD). For atom assembly we use moving tweezers superimposed on the trap beam with a polarizing beam splitter (PBS). This extra beam is deflected in the plane perpendicular to the beam propagation with a 2D acousto-optical deflector (AOD), and its focus can be displaced axially by changing the focal length of an electrically tunable lens (ETL3). The remaining electrically tunable lenses (ETL1 and ETL2) in the camera paths allow imaging of different planes along z. The inset depicts the intensity distribution of the trap light forming a bilayer array (red) and the action of the moving tweezers on an individual atom (purple). **b–d**, Intensity reconstructions of exemplary 3D patterns obtained from a collection of z-stack images taken with the diagnostics CCD camera. The regions of maximum intensity form a trefoil knot (**b**), a $5 \times 5 \times 5$ cubic array (**c**) and a C_{320} fullerene-like structure (**d**). The dimensions, L_x, L_y, L_z, of the images are the same in all the examples.

[1]Laboratoire Charles Fabry, Institut d'Optique Graduate School, CNRS, Université Paris-Saclay, Palaiseau, France. [2]These authors contributed equally: Daniel Barredo, Vincent Lienhard, Sylvain de Léséleuc. *e-mail: daniel.barredo@gmail.com

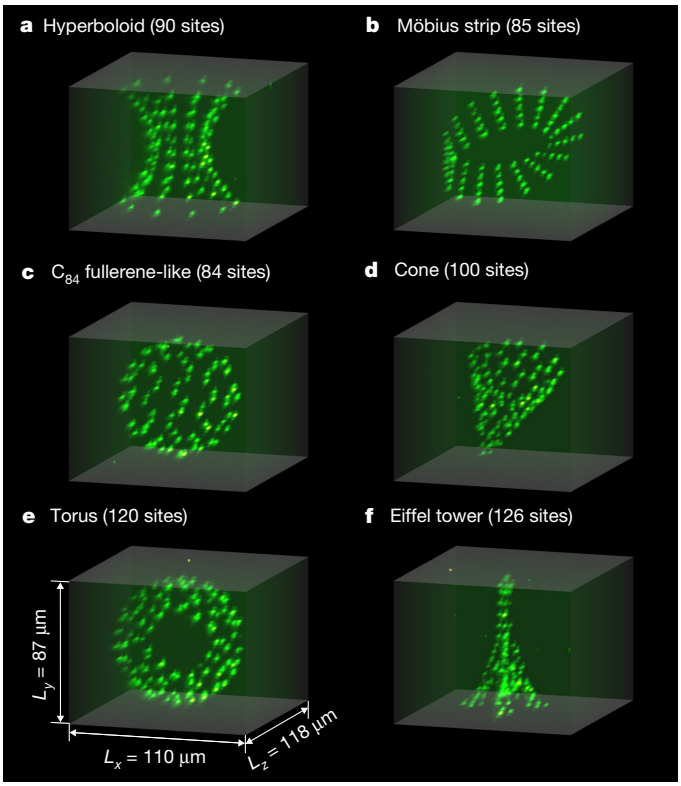


Fig. 2 | Single-atom fluorescence in 3D arrays. a–f, Maximum-intensity-projection reconstruction of the average fluorescence of single atoms loaded stochastically into exemplary arrays of traps. The x, y, z scan range of the fluorescence (L_x, L_y, L_z) is the same for all the 3D reconstructions.

optimization[14]. Rubidium-87 atoms are then loaded in the traps from a magneto-optical trap (MOT), with a final temperature of 25 μK. We detect the occupancy of each trap by collecting the fluorescence of the atoms at 780 nm with an electron-multiplying CCD camera for 50 ms. A second tunable lens (ETL2) in the imaging path is used to focus the fluorescence of different atom planes.

In Fig. 2 we show the fluorescence of single atoms trapped in various complex 3D structures, some of which are relevant, for instance, to the study of non-trivial properties of Chern insulators[19–21]. Each example is reconstructed from a series of 100 z-stack images covering an axial range of about 120 μm. With no further action, these arrays are randomly loaded with a filling fraction of about 0.5; we thus average the fluorescence signal over 300 frames to reveal the geometry of the structures.

For deterministic atom loading, we extend our two-dimensional (2D) atom-by-atom assembler[11] to 3D geometries. For that, we superimpose a second 850-nm laser beam (with $1/e^2$ radius of about 1.3 μm) on the trapping beam, which can be steered in the x–y plane using a 2D acousto-optical deflector and in the z direction by changing the focal length of a third tunable lens (ETL3). Combined with a real-time control system, the moving tweezers can perform single-atom transport with fidelities exceeding 0.993, as shown in ref. [11], and produce fully loaded arrays by using independent and sequential rearrangement of the atoms for each of the n_p planes in the 3D structures.

To explore the feasibility of plane-by-plane atom assembly, we first determine the minimal separation between layers so that each target plane can be reordered without affecting the others. To quantify this, we perform the following experiment in a 2D array containing 46 traps. We randomly load the array with single atoms and demand the atom assembler to remove all the atoms. We average over about 50 realizations and then repeat the experiment for different axial separations between the position of the moving tweezers and the trap plane. The result is shown in Fig. 3a, where we see that for separations beyond about 17 μm the effect of the moving tweezers on the atoms is negligible. This distance can be further reduced to about 14 μm by operating the moving tweezers with less power, without any degradation in the performance of the sorting process. In a complementary experiment, where we fully assembled small arrays, we also checked that the assembling efficiency is not affected by slight changes (below about 3 μm) in the exact axial position of the moving tweezers.

We now demonstrate full loading of arbitrary 3D lattices using plane-by-plane assembly. We start by creating a 3D trap array that can be decomposed in several planes normal to z. In each plane we generate approximately twice the number of traps that we need to load, so that we can easily load enough atoms to assemble the target structure. The sequence used to create fully loaded patterns (see Fig. 3b) starts by loading the MOT and monitoring the atoms entering and leaving the traps by sequentially taking a fluorescence picture for each plane. We trigger the assembler as soon as there are enough atoms in each plane to fully assemble it. We then freeze the loading by dispersing the MOT cloud and record the initial positions of the atoms by another series of z-stack images. Analysis of the images reveals which traps are filled with single atoms. We use this information to compute (in about 1 ms) the moves needed to create the fully loaded target array and perform plane-by-plane assembly by changing the z position of the moving tweezers after the assembly in each plane is completed. Finally, we detect the final 3D configuration with another series of z-stack images.

Figure 3c–h shows a gallery of fully loaded 3D atomic arrays arbitrarily arranged in space. We can create fully loaded 3D architectures with up to 72 atoms distributed in several layers with different degrees of complexity. The selected structures include simple cubic lattices (Fig. 3d), bilayers with square or graphene-like[22] arrangements (Fig. 3c, e, g), lattices with inherent geometrical frustration such as pyrochlore[23] (Fig. 3f) and lattices with cylindrical symmetry (Fig. 3h), which are suitable, for example, for studying quantum Hall physics with neutral atoms[24]. The arrays are not restricted to periodic arrangements, and the positions of the atoms can be controlled with high accuracy (<1 μm). The minimum interlayer separation that we can achieve depends on the type of underlying geometry. This is illustrated in Fig. 3e, which shows the full 3D assembly of a bilayer square lattice (with a layer separation of $d_z = 5$ μm). There, sites corresponding to the second layer are displaced by half the lattice spacing. Because traps belonging to neighbouring layers do not have the same (x, y) coordinates, there is no limitation to the minimum interlayer distance that we can produce. In both images we can observe a defocused fluorescence at intersite positions due to atoms trapped in the neighbouring layer. By contrast, whenever traps are aligned along the z axis (for example, in Fig. 3d), we set a minimum axial separation of about 17 μm to avoid any disturbance from the moving tweezers on the atoms while assembling neighbouring planes. However, for some trapping geometries this constraint can be overcome by applying a small global rotation of the 3D trap pattern around the x or y axis, so that neighbouring traps do not share the same (x, y) coordinates. The minimum interlayer spacing ultimately depends on the Rayleigh range of our trapping beam (about 5 μm) and could be further reduced, for example, by using an aspheric lens with higher numerical aperture. The range of interatomic distances that we can achieve (3– 40 μm) is suitable for implementing fast qubit gates[25] or simulating excitation transport[26] and quantum magnetism with Rydberg atoms, because interaction energies between Rydberg states at those distances are typically in the megahertz range.

To illustrate this possibility, we performed a proof-of-principle experiment with two atoms belonging to the cylindrical lattice displayed in Fig. 3h. The atoms are separated by a total distance of $R_{12} = 20$ μm ($d_x = 10$ μm, $d_z = 17$ μm); see Fig. 4. We first initialize the atoms in state $|g\rangle = |5S_{1/2}, F=2, m_F=2\rangle$, where F and m_F are the hyperfine and magnetic quantum numbers, respectively, by optical pumping in a 47-G magnetic field that defines the quantization axis and is aligned perpendicular to the internuclear axis. Then, the dipole trap is switched off and a two-photon Rydberg stimulated Raman adiabatic passage[27] excites both atoms to the $|\uparrow\rangle = |60S_{1/2}, m_j = 1/2\rangle$ Rydberg state, where m_j is the spin projection along the magnetic field direction.

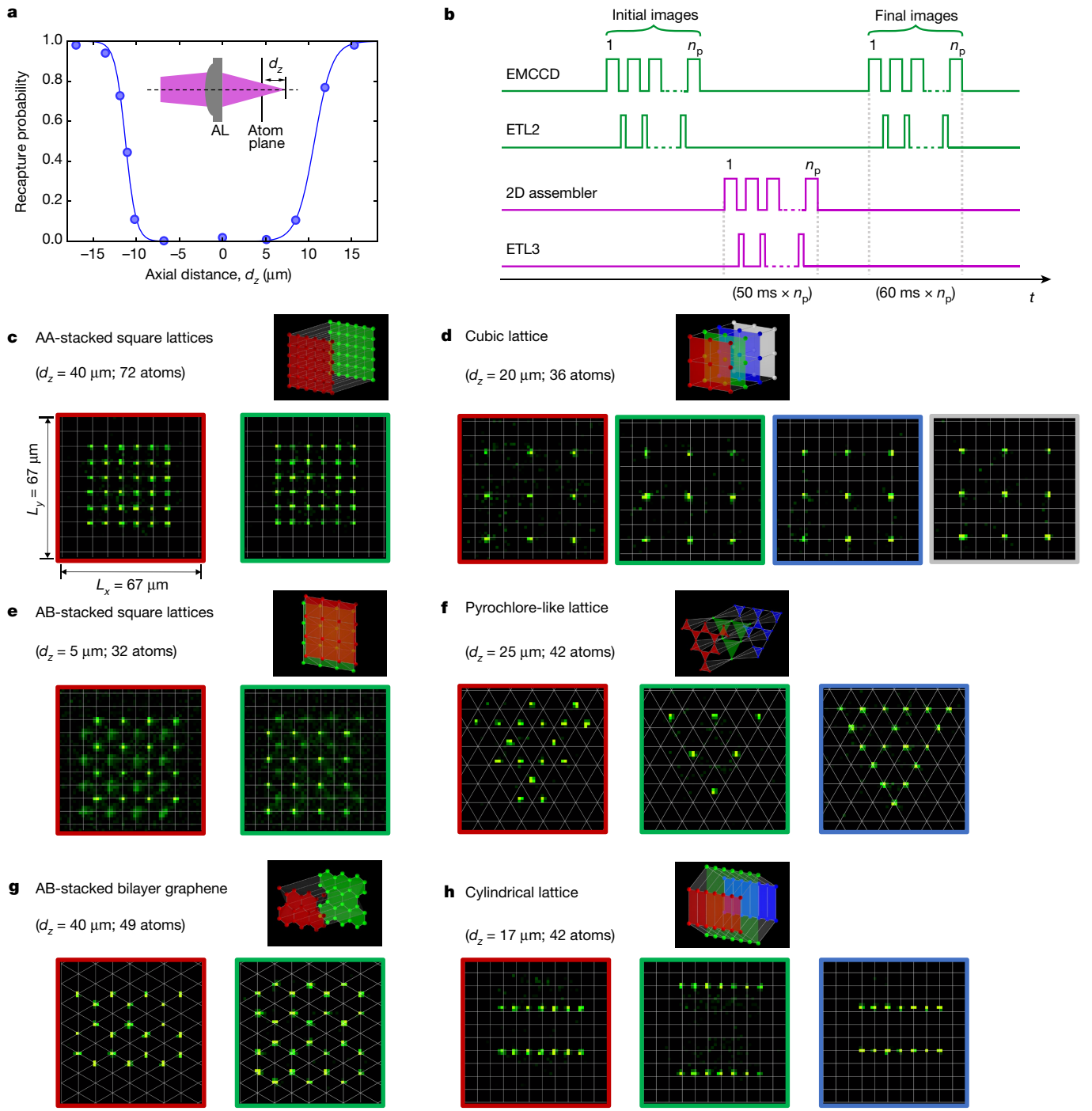


Fig. 3 | Fully loaded 3D arrays of single atoms. a, Recapture probability as a function of the axial distance between the focus of the moving tweezers and the plane of the atoms measured experimentally by trying to remove all the atoms from a 46-trap array. Error bars denote the standard error of the mean and are smaller than the symbol size. The line is a guide to the eye. **b**, Time control sequence of the experiment. We start the experiment by recording sequentially an image for each target plane. The analysis of the resulting n_p images reveals the initial position of the atoms in the traps. The 2D atom assembler, in combination with an electrically tunable lens (ETL3), arranges the atoms plane by plane. Finally, a new set of sequential images is collected to capture the result of the 3D assembly. **c**–**h**, Fully loaded arrays with arbitrary geometries. All images are single shots. The models of the 3D configurations are shown for clarity; the colours of the frames around the images encode successive atomic planes.

We further use a resonant microwave field and local addressing[28] to transfer the second atom to the $|\downarrow\rangle = |60P_{1/2}, m_j = -1/2\rangle$ state, while the first atom remains in $|\uparrow\rangle$. In these two Rydberg levels, the atoms are coupled by a direct dipole–dipole interaction with a strength of $U = C_3/R_{12}^3$, and a calculated C_3 coefficient of $C_3 = h \times 1{,}357\ \mathrm{MHz}\,\mu\mathrm{m}^3$, where h is the Planck constant. The prepared pair-state $|\uparrow\downarrow\rangle$ evolves under the XY-spin Hamiltonian $H = (C_3/R_{12}^3)\,(\sigma_1^+\,\sigma_2^- + \sigma_1^-\,\sigma_2^+)$ (where $\sigma_i^{\pm}$ denotes the Pauli matrices acting on atom $i = \{1, 2\}$) and undergoes coherent spin-exchange oscillations between $|\uparrow\downarrow\rangle$ and $|\downarrow\uparrow\rangle$ as a function of the variable interaction time, T. Finally, a de-excitation sequence projects the population in $|\uparrow\rangle$ to $|g\rangle$, but leaves the population in $|\downarrow\rangle$ unaffected. After switching the dipole trap on again, atoms in $|g\rangle$ are recaptured, while atoms in the excited state $|\downarrow\rangle$ are repelled by the trapping potential of the optical tweezers and appear as atom losses in the final fluorescence images. The outcome of this experiment is shown in Fig. 4. We observe coherent 'flip-flops' between $|\uparrow\downarrow\rangle$ and $|\downarrow\uparrow\rangle$ with a measured frequency of $2U/h = 333 \pm 5$ kHz. This value is consistent with the frequency $2U/h = 339$ kHz expected from our distance

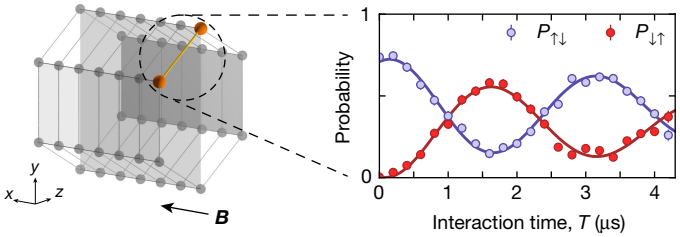


Fig. 4 | Spin-exchange dynamics between two Rydberg atoms in different *z* layers. Excitation-hopping oscillations between $|\uparrow\downarrow\rangle$ and $|\downarrow\uparrow\rangle$, observed in the populations $P_{\uparrow\downarrow}$, $P_{\downarrow\uparrow}$, driven by the dipole–dipole interaction between two Rydberg states, $|\uparrow\rangle = |60S_{1/2}, m_j = 1/2\rangle$ and $|\downarrow\rangle = |60P_{1/2}, m_j = -1/2\rangle$, at a distance of about 20 μm ($d_x = 10\,\mu m$; $d_y = 17\,\mu m$). Error bars represent the standard error of the mean and are mostly smaller than the symbol size. Solid lines are damped sine fits to the data. The direction of the magnetic field, ***B***, is indicated.

calibration ($R_{12} = 20 \pm 1\,\mu m$), which was performed by optical means. The finite contrast and the small damping of the oscillations arise from experimental imperfections (errors in state preparation and readout, residual atomic temperature), as reported in ref. [29]. This proof-of-principle experiment demonstrates the feasibility of performing quantum simulations using our defect-free 3D atomic arrays of single atoms. Excitations hopping under the influence of this Hamiltonian are equivalent to a system of hard-core bosons. The dipole–dipole interactions observed here can be further exploited to engineer Hamiltonians containing complex hopping amplitudes, which are suitable for the study of, for example, topological insulators[30].

Besides the unique tunability of the geometries that it provides, our atom-assembling procedure is highly efficient: we reach typical filling fractions of 0.95. This measured efficiency is slightly dependent on the number of planes and is mainly limited by the lifetime of the atoms in the traps (about 10 s) and the duration of the sequence (we typically need 60 ms per plane to acquire the fluorescence images and about 50 ms per plane to perform atom sorting). The repetition rate of the experiment is about 1 Hz. The number of traps and the filling fraction of the arrays could be further increased with current technology: (i) the volume of the trap array and the maximum number of traps can be extended by increasing the field of view of the aspheric lens and the laser power; (ii) the lifetime of the atoms in the traps can realistically be increased by an order of magnitude; (iii) the repetition rate of the experiment can be increased by optimizing the atom assembler[11], in particular by transferring atoms also between different planes[31]; and (iv) the initial filling fraction of the arrays could reach values exceeding 0.8 by using tailored light-assisted collisions[32,33]. Therefore, the generation of three-dimensional structures containing several hundred atoms at unit filling seems within reach, opening up many new possibilities in quantum information processing and quantum simulation with neutral atoms.

Data availability

The data presented in the figures and that support the other findings of this study are available from the corresponding author on reasonable request.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, statements of data availability and associated accession codes are available at https://doi.org/10.1038/s41586-018-0450-2

Received: 7 December 2017; Accepted: 16 July 2018;
Published online 5 September 2018.

1. Devoret, M. H. & Schoelkopf, R. J. Superconducting circuits for quantum information: an outlook. *Science* **339**, 1169–1174 (2013).
2. Veldhorst, M. et al. A two-qubit logic gate in silicon. *Nature* **526**, 410–414 (2015).
3. Kok, P. et al. Linear optical quantum computing with photonic qubits. *Rev. Mod. Phys.* **79**, 135–174 (2007).
4. Buluta, S. A. I. & Nori, F. Natural and artificial atoms for quantum computation. *Rep. Prog. Phys.* **74**, 104401 (2011).
5. Meschede, D. & Rauschenbeutel, A. Manipulating single atoms. *Adv. At. Mol. Opt. Phys.* **53**, 75–104 (2006).
6. Blatt, R. & Roos, C. F. Quantum simulations with trapped ions. *Nat. Phys.* **8**, 277–284 (2012).
7. Gross, C. & Bloch, I. Quantum simulations with ultracold atoms in optical lattices. *Science* **357**, 995–1001 (2017).
8. Weiss, D. S. & Saffman, M. Quantum computing with neutral atoms. *Phys. Today* **70**, 44–50 (2017).
9. Kuhr, S. Quantum-gas microscopes: a new tool for cold-atom quantum simulators. *Natl Sci. Rev.* **3**, 170–172 (2016).
10. Endres, M. et al. Atom-by-atom assembly of defect-free one-dimensional cold atom arrays. *Science* **354**, 1024–1027 (2016).
11. Barredo, D., de Léséleuc, S., Lienhard, V., Lahaye, T. & Browaeys, A. An atom-by-atom assembler of defect-free arbitrary two-dimensional atomic arrays. *Science* **354**, 1021–1023 (2016).
12. Nelson, K. D., Li, X. & Weiss, D. S. Imaging single atoms in a three-dimensional array. *Nat. Phys.* **3**, 556–560 (2007).
13. Wang, Y., Zhang, X., Corcovilos, T. A., Kumar, A. & Weiss, D. S. Coherent addressing of individual neutral atoms in a 3D optical lattice. *Phys. Rev. Lett.* **115**, 043003 (2015).
14. Nogrette, F. et al. Single-atom trapping in holographic 2D arrays of microtraps with arbitrary geometries. *Phys. Rev. X* **4**, 021034 (2014).
15. Kim, H. et al. In situ single-atom array synthesis using dynamic holographic optical tweezers. *Nat. Commun.* **7**, 13317 (2016).
16. Sturm, M. R., Schlosser, M., Walser, R. & Birkl, G. Quantum simulators by design: many-body physics in reconfigurable arrays of tunnel-coupled traps. *Phys. Rev. A* **95**, 063625 (2017).
17. Di Leonardo, R., Ianni, F. & Ruocco, G. Computer generation of optimal holograms for optical trap arrays. *Opt. Express* **15**, 1913–1922 (2007).
18. Wallis, J. W., Miller, T. R., Lerner, C. A. & Kleerup, E. C. Three-dimensional display in nuclear medicine. *IEEE Trans. Med. Imaging* **8**, 297–303 (1989).
19. Beugeling, W., Quelle, A. & Morais Smith, C. Nontrivial topological states on a Möbius band. *Phys. Rev. B* **89**, 235112 (2014).
20. Rüegg, A., Cof, S. & Moore, J. E. Corner states of topological fullerenes. *Phys. Rev. B* **88**, 155127 (2013).
21. Ningyuan, J., Owens, C., Sommer, A., Schuster, D. & Simon, J. Time- and site-resolved dynamics in a topological circuit. *Phys. Rev. X* **5**, 021031 (2015).
22. Castro Neto, A. H., Guinea, F., Peres, N. M. R., Novoselov, K. S. & Geim, A. K. The electronic properties of graphene. *Rev. Mod. Phys.* **81**, 109–162 (2009).
23. Bramwell, S. T. & Gringas, M. J. P. Spin ice state in frustrated magnetic pyroclore materials. *Science* **294**, 1495–1501 (2001).
24. Łacki, M. et al. Quantum Hall physics with cold atoms in cylindrical optical lattices. *Phys. Rev. A* **93**, 013604 (2016).
25. Saffman, M., Walker, T. G. & Mølmer, K. Quantum information with Rydberg atoms. *Rev. Mod. Phys.* **82**, 2313–2363 (2010).
26. Browaeys, A., Barredo, D. & Lahaye, T. Experimental investigations of dipole–dipole interactions between a few Rydberg atoms. *J. Phys. B* **49**, 152001 (2016).
27. Vitanov, N. V., Rangelov, A. A., Shore, B. W. & Bergmann, K. Stimulated Raman adiabatic passage in physics, chemistry, and beyond. *Rev. Mod. Phys.* **89**, 015006 (2017).
28. de Léséleuc, S., Barredo, D., Lienhard, V., Browaeys, A. & Lahaye, T. Optical control of the resonant dipole–dipole interaction between Rydberg atoms. *Phys. Rev. Lett.* **119**, 053202 (2017).
29. Barredo, D. et al. Coherent excitation transfer in a spin chain of three Rydberg atoms. *Phys. Rev. Lett.* **114**, 113002 (2015).
30. Weber, S. et al. Topologically protected edge states in small Rydberg systems. *Quantum Sci. Technol.* **3**, 044001 (2018).
31. Lee, W., Kim, H. & Ahn, J. Three-dimensional rearrangement of single atoms using actively controlled optical microtraps. *Opt. Express* **24**, 9816–9825 (2016).
32. Grünzweig, T., Hilliard, A., McGoven, M. & Andersen, M. Near-deterministic preparation of a single atom in an optical microtrap. *Nat. Phys.* **6**, 951–954 (2010).
33. Lester, B. J., Luick, N., Kaufman, A. M., Reynolds, C. M. & Regal, C. A. Rapid production of uniformally filled arrays of neutral atoms. *Phys. Rev. Lett.* **115**, 073003 (2015).

Acknowledgements We thank A. Läuchli for discussions. This work benefited from financial support by the EU (H2020 FET-PROACT Project RySQ), by the 'PALM' Labex (projects QUANTICA and XYLOS) and by the Région Île-de-France in the framework of DIM Nano-K.

Reviewer information *Nature* thanks W. Bakr, N. Lundblad and the other anonymous reviewer(s) for their contribution to the peer review of this work.

Author contributions D.B., V.L. and S.d.L. performed the experiments. T.L. and A.B. supervised the work. All authors made critical contributions to the work, discussed the results and contributed to the writing of the manuscript.

Competing interests The authors declare no competing interests.

Additional information
Supplementary information is available for this paper at https://doi.org/10.1038/s41586-018-0450-2.
Reprints and permissions information is available at http://www.nature.com/reprints.
Correspondence and requests for materials should be addressed to D.B.
Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

LETTER

https://doi.org/10.1038/s41586-018-0458-7

Sorting ultracold atoms in a three-dimensional optical lattice in a realization of Maxwell's demon

Aishwarya Kumar[1], Tsung-Yao Wu[1], Felipe Giraldo[1] & David S. Weiss[1]*

In 1872, Maxwell proposed his famous 'demon' thought experiment[1]. By discerning which particles in a gas are hot and which are cold, and then performing a series of reversible actions, Maxwell's demon could rearrange the particles into a manifestly lower-entropy state. This apparent violation of the second law of thermodynamics was resolved by twentieth-century theoretical work[2]: the entropy of the Universe is often increased while gathering information[3], and there is an unavoidable entropy increase associated with the demon's memory[4]. The appeal of the thought experiment has led many real experiments to be framed as demon-like. However, past experiments had no intermediate information storage[5], yielded only a small change in the system entropy[6,7] or involved systems of four or fewer particles[8–10]. Here we present an experiment that captures the full essence of Maxwell's thought experiment. We start with a randomly half-filled three-dimensional optical lattice with about 60 atoms. We make the atoms sufficiently vibrationally cold so that the initial disorder is the dominant entropy. After determining where the atoms are, we execute a series of reversible operations to create a fully filled sublattice, which is a manifestly low-entropy state. Our sorting process lowers the total entropy of the system by a factor of 2.44. This highly filled ultracold array could be used as the starting point for a neutral-atom quantum computer.

With an eye towards quantum computing and quantum simulation applications, there has been a recent boom in cold-atom sorting experiments. Atoms in a variety of arrays of dipole light traps have been impressively rearranged by moving individual traps[6,7,11,12]. The entropy associated with disordered occupancy in those cases is at most about 10% of the system entropy[6], which is dominated by vibrational excitation in the traps. Good vibrational cooling, along with well-sorted atoms, is required for cold collision-based quantum gates or quantum simulations. For Rydberg-based gates or simulations[13], although atoms are not strictly required to be vibrationally cold, colder is better. Rydberg gates using colder atoms are likely to yield higher fidelity because the atoms are less likely to change vibrational states during the gate, which can undesirably entangle atomic motion with qubit states. In general, better-localized atoms allow higher-fidelity addressing of individual atoms[14]. In blue-detuned traps, cold atoms see less light and thus scatter fewer trapping photons, which leads to longer coherence times. For instance, the coherence time in our experiment now exceeds 12 s.

Four atoms in a one-dimensional optical lattice[8] have been compacted using a method[15] similar to the one that we demonstrate here with 50 atoms in three dimensions. We note that at least about 50 qubits are needed for a quantum computer to perform a calculation that cannot be accomplished on a classical computer[16]. A three-dimensional (3D) geometry gives atoms many more nearby neighbours, which provides higher connectivity in the system. It also allows for a broad range of quantum simulations and is favourable for further scaling of the number of atoms in the system.

Our experiment proceeds as follows. We prepare a randomly 56%-filled blue-detuned 3D lattice with 4.8 μm lattice spacing[17]. By imaging polarization-gradient-cooling laser light, we determine the occupancy across the lattice with an error of 10^{-3} per site in 800 ms (ref. [17]). Projection sideband cooling[18] puts 89% of the caesium atoms into their vibrational ground states and >99.7% of them in the $|F=4, m_F=-4\rangle$ hyperfine ground state, where F and m_F are the hyperfine and magnetic quantum numbers, respectively. We then combine the ability to address atoms at individual sites (by using crossed laser beams and microwaves to make site-dependent state changes[19]) with the ability to make state-dependent lattice translations (by rotating the lattice beam polarizations[20]). Starting from a given 3D occupancy map we devise a sequence of operations to fill up either a $5 \times 5 \times 2$ or a $4 \times 4 \times 3$ sublattice.

We can target any site in a $5 \times 5 \times 5$ lattice by using a pair of focused addressing beams intersecting at a right angle[14,19]. Targeting proceeds as in our previous demonstration of high-fidelity single-qubit gates[14], but the magnetic sublevels are different and in this case we are unconcerned with long-term quantum coherence. The addressing beams shift the ($|F=4, m_F=-4\rangle$ to $|F=3, m_F=-3\rangle$ resonance by around 50 kHz, which allows us to drive the associated microwave transition using an adiabatic fast-passage pulse (see Methods for details) that transfers only the target atom. An atom making the transition from $m_F=-4$ to $m_F=-3$ moves from the 'stationary' to the 'motion' state.

The linear polarizations of the two beams that create the lattice in a given direction are initially aligned, so the two states are trapped nearly identically. When the polarization of one of the lattice beams is rotated (using two electro-optic modulators and a $\lambda/4$ plate, where λ is the wavelength), the optical lattices for the two states move in opposite directions (see Fig. 1a). After rotating the polarization by π, we optically pump the atoms back to the stationary state and rotate the polarization back. The net effect of this sequence is that atoms that start in the stationary state move but return to the same place, while atoms that start in the motion state are shifted by one lattice site.

The sorting algorithm for compacting atoms in the lattice was proposed in previous work[15,21]; we have slightly modified it to allow the filling of any continuous sublattice (see Methods). The general idea is to first perform a series of balancing steps in the x and y directions so that every row in the z direction has the required number of atoms to fill a desired number of planes. Then, a series of compaction steps in the z direction moves atoms to fill the planes of the target sublattice (Fig. 1b). For example, to fill a $5 \times 5 \times 2$ sublattice from a half-filled $5 \times 5 \times 5$ lattice, atoms are first 'balanced' in the x and y directions so that every row in the z direction has at least two atoms. Parallel z-motion steps then move the atoms to the desired planes. After sorting, we reimage the atoms and repeat the procedure to correct any errors. The ability to know exactly where the vacancies are is an advantage of this approach to filling a lattice compared to implementing a superfluid–Mott insulator transition[22], where residual occupancy errors are unknown.

Figure 2 shows two implementations of this algorithm, in which target sublattices were completely filled after two sorts. In general, starting with at least half the lattice sites filled in a $5 \times 5 \times 5$ array, three sorts leave us with an average filling fraction of 0.97 for $5 \times 5 \times 2$ and 0.95 for $4 \times 4 \times 3$. We achieve the perfect filling shown in Fig. 2f and Fig. 2c 32% and 27% of the time, respectively. For the first sort, the

[1]Department of Physics, The Pennsylvania State University, University Park, PA, USA. *e-mail: dsweiss@phys.psu.edu

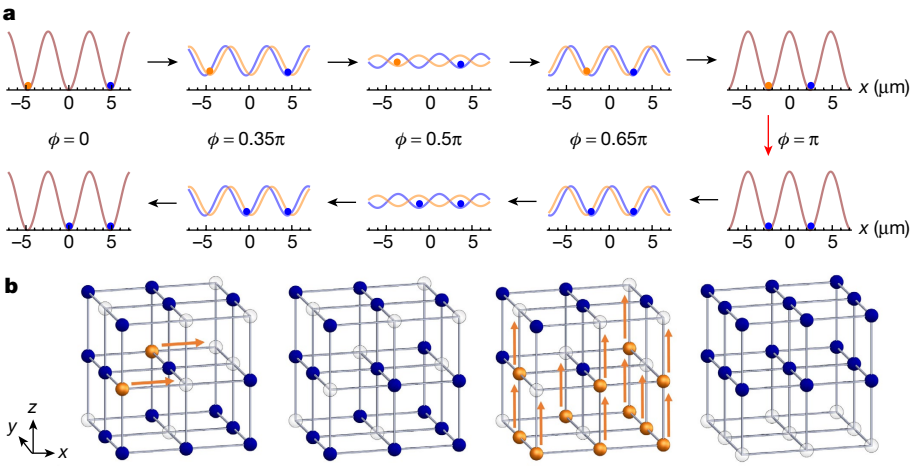


Fig. 1 | Motion steps and sorting algorithm. a, Motion steps used to fill a vacancy in a given direction. The curves show the lattice potential as a function of position for the 'motion' state (orange curve) and the 'stationary' state (blue curve). Brown curves indicate overlapping potentials. The arrows denote the direction of a time series in which the angle (ϕ) between the polarizations of the two lattice beams is adiabatically ramped to π and back to 0. The atom to be moved is transferred to the motion state (orange circle) using targeted addressing at the beginning of the time series. As the polarization of one of the lattice beams is rotated, the atoms in the motion state and the stationary state (blue circle) move in opposite directions, settling half a lattice spacing away from their original positions when $\phi = \pi$. The atom in the motion state is then optically pumped to the stationary state (illustrated by the red arrow). As the polarization is rotated back, both atoms move in the same direction, with the atom that started in the stationary state returning to its original position and the atom that started in the motion state moving by one lattice site. **b**, Simplified illustration of two parts of the sorting algorithm in a $3 \times 3 \times 3$ lattice. Orange and blue circles are as in **a**; empty circles denote empty sites. The first motion step 'balances' the array so that every z row has exactly two atoms. The second motion step 'compacts' atoms into two planes.

average number of motion steps is 6.4 (5.6) and the average number of addressing operations was 38 (62) for filling a $5 \times 5 \times 2$ ($4 \times 4 \times 3$) sublattice. Each sort takes about 190 ms on average. Figure 3 shows the filling fraction as a function of the number of sorts. These numbers match well with Monte Carlo simulations that consider measured sources of error (see Methods). A major source of error for atoms in both the motion and stationary states is spontaneous emission from the lattice. The spontaneous emission rate is significantly higher (17 times on average) during a motion step because the lattice intensity is not zero at the trap minima during the motion (see Fig. 1a). When an atom spontaneously emits a photon and changes hyperfine state, it becomes anti-trapped and is lost. The measured average loss per motion step is about 4×10^{-3}. Another source of error is imperfect transfer of atoms from the stationary state to the motion state. Our measured transfer fidelity is 0.986, limited by a combination of imperfect addressing beam shape, pointing noise of the addressing beams and magnetic field fluctuations. This error can cause two atoms to end up in the same lattice site, both of which are lost during imaging. The number of sorts that can be performed to fill errors is eventually limited by the 92-s vacuum lifetime and by double-atom loss. Optical pumping leads to a modest amount of heating, exciting about 7% of the population from the 3D vibrational ground state per motion step. Were we to replace the more convenient optical pumping with targeted addressing, this number would be reduced to 0.6%.

After sorting and a final round of projection cooling, we measure the vibrational sidebands to determine the final ground-state occupation, as shown in Fig. 4. Projection sideband cooling (see Methods) leads to ground-state occupation probabilities of 0.949(7), 0.954(6) and 0.985(1) in the x, y and z directions, respectively, which implies 89% occupation of the 3D vibrational ground state. The state is not thermal, but most of the population is in the lowest three levels. We calculate that the vibrational entropy for this state is about $0.59k_B$ per particle, where k_B is the Boltzmann constant.

The configurational entropy is given by[23]

$$S = \frac{1}{\overline{n}}\left[\overline{n}\ln\left(\frac{1}{\overline{n}}\right) + (1-\overline{n})\ln\left(\frac{1}{1-\overline{n}}\right)\right] \quad (1)$$

where $\overline{n}$ is the filling fraction. The solid blue line in Fig. 3 shows the configurational entropy as a function of the number of sorts, and the dotted line shows the vibrational entropy after projection cooling. Sorting reduces the configurational entropy by a factor of 8 and the total entropy by a factor of 2.44. The final total entropy per particle is $0.75k_B$.

The number of required motion steps scales as $N^{1/3}$, where N is the number of atoms to be sorted[15,21]. Similar scaling for state flipping could be obtained if the addressing beams were generated holographically; such a versatile 3D light pattern would allow many atoms to be state-flipped with microwaves simultaneously. Monte Carlo simulations using our current sequential addressing scheme and error rate show that starting from a half-filled $10 \times 10 \times 10$ lattice, $10 \times 10 \times 4$ and $7 \times 7 \times 7$ sublattices can be filled to a filling fraction of about 0.93. The error due to the motion could be reduced by further detuning the lattice light. Tripling the lattice detuning would decrease the spontaneous emission rate by a factor of 9 and the lattice depth by a factor of 3. Although the resulting lower trap frequency would require that we move atoms three times more slowly, the total spontaneous emission per motion step would be reduced by a factor of three, which would improve the filling fraction to approximately 0.975 for about 400 sorted atoms. It should be possible to improve microwave transfer errors by an order of magnitude by improving magnetic field stabilization and adapting our phase gate[14], which is insensitive to addressing-beam intensity fluctuations. Because atoms in a 3D lattice geometry have many near neighbours, small known filling errors can be readily incorporated into the design of any quantum computation.

We now more fully discuss our characterization of this experiment as the first, to our knowledge, to capture the full essence of Maxwell's demon on a large array of particles. Any process that involves selectively acting on particles differently depending on their momentum, energy or internal state, like all laser-cooling methods[5,24,25], evokes an aspect of Maxwell's demon, who sorted particles based on velocity. However, when there is no stored information, such mechanisms differ in spirit from Maxwell's demon and other thought-experiment demons[26]. When the entropy increase of the outside world is built into the cooling cycle, carried away by lost particles or scattered light, there is no trace of the

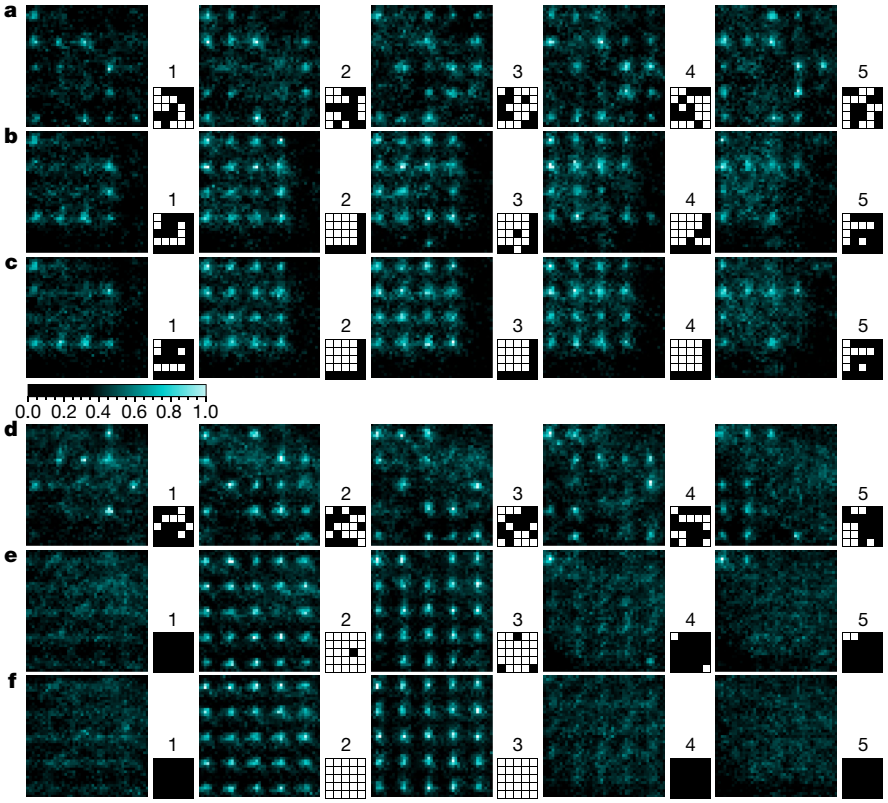


Fig. 2 | Perfect filling of 4 × 4 × 3 and 5 × 5 × 2 sublattices. a–f, The five images in each row correspond to the five lattice planes (labelled 1–5). The colour map shows intensity. We have applied contrast enhancement (a threshold of ~35% of the peak intensity) to make empty sites more obvious in the figure. The associated grid patterns are real-time occupancy maps, generated by processing the five images. The two sets of images (**a**–**c** and **d**–**f**) are from two different experimental implementations. **a**, An initial, unsorted atom distribution. The occupancy maps are used as the basis of a series of site-selective state flips and state-selective translations that execute our sorting algorithm. **b**, Result after one sorting sequence with the goal of filling a 4 × 4 × 3 sublattice in planes 2–4. There are three errors after this sort (one in plane 3, two in plane 4). **c**, Result after a second sorting sequence starting from the distribution in **b**. The sorting goal has been reached. Atoms outside the target sublattice can be kept as spares, or they can be selectively state-flipped and removed by a resonant clearing beam. **d**, Another initial, unsorted atom distribution. **e**, The result after one sorting sequence with the goal of filling a 5 × 5 × 2 sublattice in planes 2 and 3. There are four errors after this sort (one in plane 2, three in plane 3). **f**, The result after a second sorting sequence starting from the distribution in **e**. The sorting goal has been reached. The absence of spare atoms in **f** is coincidental.

theoretical paradox that twentieth-century information theory worked to resolve[2].

By contrast, our experiment is conceptually similar to Maxwell's thought experiment. We increase the entropy of the outside world in the process of determining site occupancy. At the same time, the configurational entropy goes to zero because there is only one state with that particular configuration. The stored occupancy information is then used as a guide to the execution of reversible operations that leave the system in a manifestly low-entropy state. Of course, that is also true for any sorting operation, as when checkers are arrayed on a board. The difference here is that most of the initial entropy of our system is in the initial configurational disorder, so that by measuring and sorting we considerably reduce the total system entropy. Maxwell's demon collected information and acted on one particle at a time. By contrast, our demon obtains an occupancy map of the whole system, so that it can map out a plan to act on all the particles in parallel.

Maxwell visualized work being extracted from the reconfigured system by using the demon-imposed temperature gradient to drive a heat engine. Work can probably not be extracted in our experiment, but the fact that the overall system entropy is reduced means that trap changes that affect all atoms in the same way can create a much colder gas. For instance, the experiment would pass the $1.24k_B$ entropy-per-particle threshold below which there would be a Bose–Einstein condensate if the lattice were adiabatically shut off and the atoms were

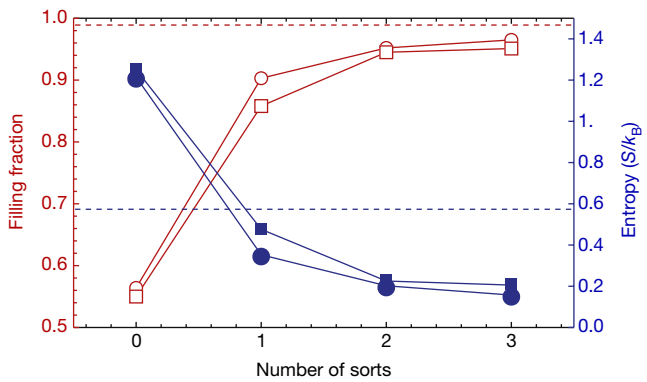


Fig. 3 | Filling fraction and entropy. The empty red symbols show the filling fraction as a function of the number of sorts for 5 × 5 × 2 (circles) and 4 × 4 × 3 (squares) target sublattices. The circles (squares) show results based on 85 (48) experimental implementations. The red horizontal dashed line is the limit associated with loss from collisions with background gas atoms during the 1 s required to image and sort. The solid blue symbols show the configurational entropy as a function of the number of sorts for 5 × 5 × 2 (circles) and 4 × 4 × 3 (squares) target sublattices. The total entropy at the beginning and at the end is the sum of the vibrational entropy (blue horizontal dotted line) and the configurational entropy; sorting reduces it by a factor of 2.44. The 1σ error bars are smaller than the size of the symbols.

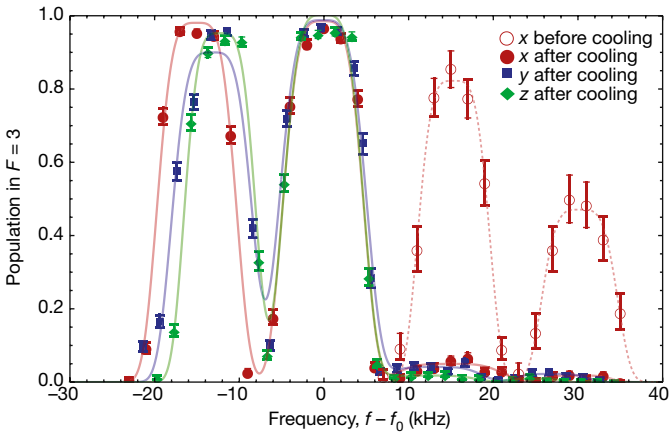


Fig. 4 | Microwave spectra showing the results of projection sideband cooling[18]**.** Atoms start in $|F=4, m_F=-4\rangle$, and the curves show the number of atoms that make an adiabatic fast-passage transition from the $|F=4, m_F=-4\rangle$ to the $|F=3, m_F=-3\rangle$ state because atoms that remain in $|F=4, m_F=-4\rangle$ are cleared by a resonant beam before detection. Slightly rotating the polarization of one lattice beam of the x, y or z lattice beam pairs leads to non-zero projections between different vibrational states (vibrational quantum number ν) in that direction. Nearly all atoms make the transition at the peaks centred at zero, where the frequency is $f=f_0\approx 9.19130$ GHz, which corresponds to no change of vibrational level. Nearly all atoms make the transition at the lower-frequency peaks (at $f-f_0\approx -15$ kHz), which corresponds to the vibrational quantum number increasing by one ($\Delta\nu=+1$). Atoms make the transition at the higher-frequency peaks ($\Delta\nu=-1$, at $f-f_0\approx 15$ kHz) only if there is a lower vibrational level available, which is true for all atoms except those in the vibrational ground state. The heights of the $\Delta\nu=-1$ peaks thus provide a measure of the atoms that are not in the vibrational ground state. The empty maroon circles show the $\Delta\nu=-1$ and $\Delta\nu=-2$ sidebands in the x direction before projection sideband cooling. The sidebands for the y and z directions are similar (not shown). The solid maroon circles, solid blue squares and solid green diamonds show the spectra for the x, y and z directions, respectively, after projection sideband cooling. Each data point in the 'before cooling' dataset is obtained from about 60 atoms and in the 'after cooling' datasets from about 400 atoms. All error bars represent one standard deviation. The maroon, blue and green solid lines are sums of four fitted super-Gaussians of order 4 for the x, y and z directions, respectively. We infer from the large suppressions of the $\Delta\nu=-1$ and $\Delta\nu=-2$ sidebands in this figure that the 3D vibrational ground state is occupied 88.9(9)% of the time.

left in a 3D box potential[23,27]. The adiabatic timescale for making this approximate Mott-insulator-to-superfluid transition is far too large given our current lattice spacing, but it could be accomplished if the lattices were made in an accordion configuration[28], in which the angle between the beam pairs could be dynamically changed to reduce the lattice constant by a factor of five. That would present a third path to quantum degeneracy of cold atoms, joining many evaporative-cooling experiments and one laser-cooling experiment[25,29].

Maxwell's thought experiment led to a deep understanding of the relationship between entropy and information. The experiment that we present here has several practical applications. It prepares a favourable initial state for a neutral-atom quantum computer with one atom at nearly every 3D lattice site, each cooled near its vibrational ground state. The cold array minimizes many potential errors in Rydberg-gate-based quantum computations. The 3D optical lattice allows entanglement with many near neighbours and provides favourable scaling, often as $N^{1/3}$ or $N^{2/3}$, to minimize computation time and the laser power requirements. If we can further improve cooling—for instance, by temporarily transferring the atoms to a lattice with smaller detuning, where the atoms are trapped more deeply in the Lamb–Dicke limit—we might be able to create large-scale entanglement through cold-collision gates[30,31] and thus ultimately implement one-way quantum computation[32]. Our sorted array could also be used for a variety of Rydberg-based quantum simulations with different geometries, dimensionalities and anisotropy of interactions. For instance, the simulations of Ising-like Hamiltonians that have recently been implemented in one- and two-dimensional (2D) tweezer and microtrap arrays[16,33] could be extended to three dimensions in our optical lattice. Our demonstrated coherent site-selective control[14] allows the possible implementation of a universal quantum simulator, which might be used to implement Kitaev's toric code in 2D sublattices and in 2D or 3D lattice gauge theories[34].

Online content

Any methods, additional references, Nature Research reporting summaries, source data, statements of data availability and associated accession codes are available at https://doi.org/10.1038/s41586-018-0458-7.

Received: 15 January 2018; Accepted: 16 July 2018;
Published online 5 September 2018.

1. Maxwell, J. C. *Theory of Heat* (Longmans, Green and Co., London, 1871).
2. Leff, H. S. & Rex, A. F. *Maxwell's Demon: Entropy, Information, Computing* (Princeton University Press, Princeton, 1990).
3. Brillouin, L. Maxwell's demon cannot operate – information and entropy. 1. *J. Appl. Phys.* **22**, 334–337 (1951).
4. Landauer, R. Irreversibility and heat generation in the computing process. *IBM J. Res. Dev.* **5**, 183–191 (1961).
5. Price, G. N., Bannerman, S. T., Viering, K., Narevicius, E. & Raizen, M. G. Single-photon atomic cooling. *Phys. Rev. Lett.* **100**, 093004 (2008).
6. Barredo, D., de Leseleuc, S., Lienhard, V., Lahaye, T. & Browaeys, A. An atom-by-atom assembler of defect-free arbitrary two-dimensional atomic arrays. *Science* **354**, 1021–1023 (2016).
7. Endres, M. et al. Atom-by-atom assembly of defect-free one-dimensional cold atom arrays. *Science* **354**, 1024–1027 (2016).
8. Robens, C. et al. Low-entropy states of neutral atoms in polarization-synthesized optical lattices. *Phys. Rev. Lett.* **118**, 065302 (2017).
9. Lester, B. J., Luick, N., Kaufman, A. M., Reynolds, C. M. & Regal, C. A. Rapid production of uniformly filled arrays of neutral atoms. *Phys. Rev. Lett.* **115**, 073003 (2015).
10. Strasberg, P., Schaller, G., Brandes, T. & Esposito, M. Thermodynamics of a physical model implementing a Maxwell demon. *Phys. Rev. Lett.* **110**, 040601 (2013).
11. Barredo, D., Lienhard, V., de Léséleuc, S., Lahaye, T. & Browaeys, A. Synthetic three-dimensional atomic structures assembled atom by atom. *Nature* http://dx.doi.org/10.1038/s41586-018-0450-2 (2018).
12. Kim, H. et al. In situ single-atom array synthesis using dynamic holographic optical tweezers. *Nat. Commun.* **7**, 13317 (2016).
13. Saffman, M. Quantum computing with atomic qubits and Rydberg interactions: progress and challenges. *J. Phys. B* **49**, 202001 (2016).
14. Wang, Y., Kumar, A., Wu, T. Y. & Weiss, D. S. Single-qubit gates based on targeted phase shifts in a 3D neutral atom array. *Science* **352**, 1562–1565 (2016).
15. Weiss, D. S. et al. Another way to approach zero entropy for a finite system of atoms. *Phys. Rev. A* **70**, 040302 (2004).
16. Bernien, H. et al. Probing many-body dynamics on a 51-atom quantum simulator. *Nature* **551**, 579–584 (2017).
17. Nelson, K. D., Li, X. & Weiss, D. S. Imaging single atoms in a three-dimensional array. *Nat. Phys.* **3**, 556–560 (2007).
18. Li, X., Corcovilos, T. A., Wang, Y. & Weiss, D. S. 3D projection sideband cooling. *Phys. Rev. Lett.* **108**, 103001 (2012).
19. Wang, Y., Zhang, X. L., Corcovilos, T. A., Kumar, A. & Weiss, D. S. Coherent addressing of individual neutral atoms in a 3D optical lattice. *Phys. Rev. Lett.* **115**, 043003 (2015).
20. Deutsch, I. H. & Jessen, P. S. Quantum-state control in optical lattices. *Phys. Rev. A* **57**, 1972–1986 (1998).
21. Vala, J. et al. Perfect pattern formation of neutral atoms in an addressable optical lattice. *Phys. Rev. A* **71**, 032324 (2005).
22. Jaksch, D., Bruder, C., Cirac, J. I., Gardiner, C. W. & Zoller, P. Cold bosonic atoms in optical lattices. *Phys. Rev. Lett.* **81**, 3108–3111 (1998).
23. Olshanii, M. & Weiss, D. Producing Bose–Einstein condensates using optical lattices. *Phys. Rev. Lett.* **89**, 090404 (2002).
24. Chu, S., Hollberg, L., Bjorkholm, J. E., Cable, A. & Ashkin, A. Three-dimensional viscous confinement and cooling of atoms by resonance radiation pressure. *Phys. Rev. Lett.* **55**, 48–51 (1985).
25. Anderson, M. H., Ensher, J. R., Matthews, M. R., Wieman, C. E. & Cornell, E. A. Observation of Bose–Einstein condensation in a dilute atomic vapor. *Science* **269**, 198–201 (1995).
26. Szilard, L. Über die Entropieverminderung in einem thermodynamischen System bei Eingriffen intelligenter Wesen. *Z. Phys.* **53**, 840–856 (1929).
27. Gaunt, A. L., Schmidutz, T. F., Gotlibovych, I., Smith, R. P. & Hadzibabic, Z. Bose–Einstein condensation of atoms in a uniform potential. *Phys. Rev. Lett.* **110**, 200406 (2013).
28. Williams, R. A. et al. Dynamic optical lattices: two-dimensional rotating and accordion lattices for ultracold atoms. *Opt. Express* **16**, 16977–16983 (2008).
29. Hu, J. Z. et al. Creation of a Bose-condensed gas of Rb-87 by laser cooling. *Science* **358**, 1078–1080 (2017).

30. Jaksch, D., Briegel, H. J., Cirac, J. I., Gardiner, C. W. & Zoller, P. Entanglement of atoms via cold controlled collisions. *Phys. Rev. Lett.* **82**, 1975–1978 (1999).
31. Kaufman, A. M. et al. Entangling two transportable neutral atoms via local spin exchange. *Nature* **527**, 208–211 (2015).
32. Raussendorf, R. & Briegel, H. J. A one-way quantum computer. *Phys. Rev. Lett.* **86**, 5188–5191 (2001).
33. Labuhn, H. et al. Tunable two-dimensional arrays of single Rydberg atoms for realizing quantum Ising models. *Nature* **534**, 667–670 (2016).
34. Weimer, H., Muller, M., Lesanovsky, I., Zoller, P. & Buchler, H. P. A Rydberg quantum simulator. *Nat. Phys.* **6**, 382–388 (2010).

Acknowledgements This work was supported by the US National Science Foundation through grant PHY-1520976.

Author contributions All authors contributed to the design, execution and analysis of the experiment and the writing of the manuscript. A.K., T.-Y.W. and F.G. collected all the data.

Competing interests The authors declare no competing interests.

Additional information
Extended data is available for this paper at https://doi.org/10.1038/s41586-018-0458-7.
Reprints and permissions information is available at http://www.nature.com/reprints.
Correspondence and requests for materials should be addressed to D.S.W.
Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

METHODS

Apparatus. We load atoms from a Magneto-Optical Trap to a 3D optical lattice formed by three pairs of 75-μm-waist, 838.95-nm laser beams. Each lattice beam has a power of 250 mW, giving a lattice depth of 190 μK at the central lattice site. The two beams in each pair cross each other at 10°, yielding a lattice spacing of 4.9 μm. Two pairs are frequency-shifted relative to the third by +30 MHz and −175 MHz to prevent mutual interference among the lattice pairs. One beam in each pair has two electro-optic modulators in its path, aligned so that their axes are at 45° relative to the incoming polarization, followed by a λ/4 wave-plate aligned with the incoming polarization. As the voltage on the electro-optic modulators is increased, the polarization of this beam rotates. The angle of rotation is π when the half-wave voltage is applied to both electro-optic modulators.

Projection sideband cooling. Projection sideband cooling has been described in detail in a previous paper[18]. We give a brief overview here. Projection sideband cooling is similar to other sideband-cooling techniques, except that microwave photons, which have a very small momentum compared to the optical photons that are usually used, drive the vibrational-state-changing transitions. To accomplish this, for a given lattice direction, say x, the polarization of one of the lattice beams in the pair is rotated slightly, which displaces the traps experienced by atoms in the $|F=4, m_F=-4\rangle$ and $|F=3, m_F=-3\rangle$ states slightly relative to each other (the same effect that we use to state-selectively translate atoms during our sorting operations). Therefore, all the vibrational wavefunctions associated with one magnetic sublevel have non-zero projections onto the wavefunctions of the other sublevel. There are consequently non-zero matrix elements for vibration-state-changing microwave transitions.

Projection cooling proceeds as follows. All the atoms are prepared in $|F=4, m_F=-4\rangle$ via optical pumping. The x lattice-beam polarization is rotated, and then a $\Delta\nu_x=-2$ microwave adiabatic fast passage (AFP) pulse drives the transition $|F=4, m_F=-4\rangle$ to $|F=3, m_F=-3\rangle$, followed by a modified polarization rotation and then a $\Delta\nu_x=1$ microwave pulse from $|F=3, m_F=-3\rangle$ to $|F=4, m_F=-4\rangle$ (which is a $\Delta\nu=-1$ pulse for atoms that start in $|F=4, m_F=-4\rangle$). The AFP pulses work well regardless of the initial vibrational state. The two AFP pulses lower ν_x by 1 for all the atoms except those initially in $\nu_x=0$, which make no transitions. All the atoms for which both AFP pulses were successful end in $|F=4, m_F=-4\rangle$, except those that started in $\nu_x=1$. The polarization is rotated back, and an optical pumping pulse resets all the atoms to $|F=4, m_F=-4\rangle$. These steps are then repeated for the y and z lattice beams, and then the sequence is repeated 50 times. The whole cooling sequence takes about 1 s. The $\Delta\nu=-2\rightarrow\Delta\nu=1$ sequences minimize the number of times that an atom has to be optical pumped, which is a particular advantage for cooling atoms from high vibrational states. Sequentially stepping through the Cartesian directions optimizes the final cooling steps.

We have improved the performance of our previously demonstrated projection sideband cooling method[18] considerably, from 76% to 89% occupancy of the ground vibrational state. This improvement results from two changes. First, we increased the fidelity of the $\Delta\nu=-2$ microwave pulse, where $\Delta\nu$ is the microwave-driven change in vibrational level, by separately optimizing the lattice displacement for $\Delta\nu=-2$ and $\Delta\nu=-1$. Second, we improved the quality of the optical-pumping-light polarization at the atoms by a factor of 5.

Implementation of a motion step. Extended Data Fig. 1 illustrates our timing sequence for one motion step. Before any motion, atoms are optically pumped to the $|F=4, m_F=-4\rangle$ state (not shown). The atoms to be moved are transferred to the $|F=3, m_F=-3\rangle$ state sequentially. The addressing lasers, directed by micro-mechanical electronic systems mirrors, cross at a target atom in the 3D array, causing an a.c. Stark shift on its resonance frequency between $|F=4, m_F=-4\rangle$ and $|F=3, m_F=-3\rangle$ by −50 kHz with respect to the atoms that are not in the path of either addressing laser. The addressing laser powers are ramped up over 40 μs, after which we wait for another 110 μs for our intensity lock to settle. We drive the transition in the target atoms with a 3-ms-long AFP microwave pulse, which involves a 12-kHz frequency sweep. The crosstalk is less than 3×10^{-3}.

To initiate motion, the polarization of one of the lattice beams is linearly rotated by π over 3 ms by ramping the voltages on the electro-optic modulators. The atoms in $|F=3, m_F=-3\rangle$ are then optically pumped to $|F=4, m_F=-4\rangle$ in 0.2 ms (with an intensity of 4 mW cm^{-2} and detuning of −7.5 MHz on the $F=3$ to $F'=4$ transition, and 0.5 mW cm^{-2} and 7.5 MHz on the $F=4$ to $F'=4$ transition). The voltages are then ramped back to zero. A final optical pumping step over 0.25 ms ensures that all atoms are back to $|F=4, m_F=-4\rangle$ for the next motion step.

Measuring state-flip fidelity. To measure the efficiency of our addressing scheme, we take an occupancy map, apply projection sideband cooling to the atoms and optically pump them to the $|F=4, m_F=-4\rangle$ state. We then sequentially flip the state of all the atoms within a $5\times5\times5$ region to $|F=3, m_F=-3\rangle$ using targeted addressing. Then, another laser beam resonant with the transition from $|F=4\rangle$ to $|F'=5\rangle$ pushes away the atoms that were left in the $|F=4, m_F=-4\rangle$ state. A new occupancy map is then generated to identify the atoms that were successfully transferred to $|F=3, m_F=-3\rangle$. Averaging over 50 implementations, we measure a state-flip fidelity of 0.986(5). However, the addressing laser beam drifts slowly once aligned, which can decrease the state-flip fidelity by about 0.02 after about 100 sorting operations.

Measuring motion fidelities. Motion errors can occur when atoms spontaneously emit lattice light. An atom is usually lost during motion if light scattering leaves it in the anti-trapped state. Occasionally the atom site-hops, if it stays trapped but follows the 'wrong' lattice potential. We measured the motion fidelities for atoms in $|F=4, m_F=-4\rangle$ and $|F=3, m_F=-3\rangle$ separately (see Extended Data Fig. 2). Atoms were first projection-sideband-cooled and optically pumped to $|F=4, m_F=-4\rangle$. To find the cumulative effect of making $2N$ motion steps in a given direction, we ramped up to the half-wave voltage, $V_{\lambda/2}$, of the electro-optic modulators, then ramped down to $-V_{\lambda/2}$, and repeated the process N times. Because no optical pumping or state flips were applied during the motions, all the atoms moved back and forth by one lattice spacing around their initial positions. By comparing the occupancy maps before and after these motions steps, we can identify the percentage of atoms that successfully return to their initial positions, which we call the motion fidelity. For motion in $|F=3, m_F=-3\rangle$, the sequence is the same except that after the atoms are optically pumped to $|F=4, m_F=-4\rangle$, a global microwave pulse is applied to flip the state of all atoms to $|F=3, m_F=-3\rangle$ before executing the motions. Each data point in Extended Data Fig. 2 is averaged over 10 sorting operations and corrected for the loss due to collisions with background gas atoms. A linear fit gives the fidelities per motion step. For atoms in the $|F=4, m_F=-4\rangle$ state and motion in the lattice directions, the fidelities are {0.9951(6), 0.9982(6), 0.9962(4)}, where the errors refer to one standard deviation. The corresponding fidelities for atoms in the $|F=3, m_F=-3\rangle$ state are {0.9956(4), 0.9961(10), 0.9956(1)}. The calculated probability of spontaneous emission for an atom in the vibrational ground state during a motion step is 3.5×10^{-3}.

Sorting algorithm. We have generalized the sorting algorithm for any initial $N\times N\times N$ lattice and any final $i\times j\times k$ sublattice. If $i=j=N$, then only balancing and compaction steps are needed. If $i, j<N$, then extra motion steps in x and y are added to move as many atoms as possible into an $i\times j\times N$ sublattice from 'outside' before balancing and compaction ('outside' means the full lattice minus the target sublattice). For example, to fill a $4\times4\times3$ sublattice, as many atoms as possible are first moved into a $4\times4\times5$ region in two motion steps, one in x and one in y, from outer y–z and x–z planes of the lattice. Balancing and compaction are then applied to a $4\times4\times5$ lattice rather than a $5\times5\times5$ lattice. The simulations that we describe below suggest that even though this procedure does not always empty the outside planes, there are always enough atoms to fill a $4\times4\times3$ sublattice when starting from a 50% filled $5\times5\times5$ lattice.

The steps for balancing an $i\times j\times N$ lattice to fill an $i\times j\times k$ sublattice are roughly as follows:

1. If this is the first iteration, choose a dividing plane, P, to be an x–z plane. Otherwise, choose the dividing plane to be perpendicular (either x–z or y–z) to the previous iteration. Choose P to divide the lattice into two parts, S_1 and S_2, that are as similar as possible (that is, a difference of one plane between S_1 and S_2 is permitted if the lattice dimension is odd).
2. If the number of z rows in S_1 (S_2) is n (m), the required number of atoms in S_1 (S_2) is $k\times n$ ($k\times m$). Move atoms between the two sublattices until they each have at least the required number of atoms.
3. Repeat these steps for S_1 and S_2 separately, stopping when each of them is just a single z row.

Balancing guarantees that there are k atoms in each of the $i\times j$ z rows. These atoms are then moved in the z direction ('compaction') in parallel to fill the desired k planes, usually in the middle of the accessible lattice. The algorithm minimizes the number of motion steps.

The sorting algorithm can probably be improved by replacing the initial steps to empty the outer x–z and y–z planes by a more optimal algorithm. For instance, the first sort could be modified to distribute the extra atoms evenly and thus reduce the number of correction steps.

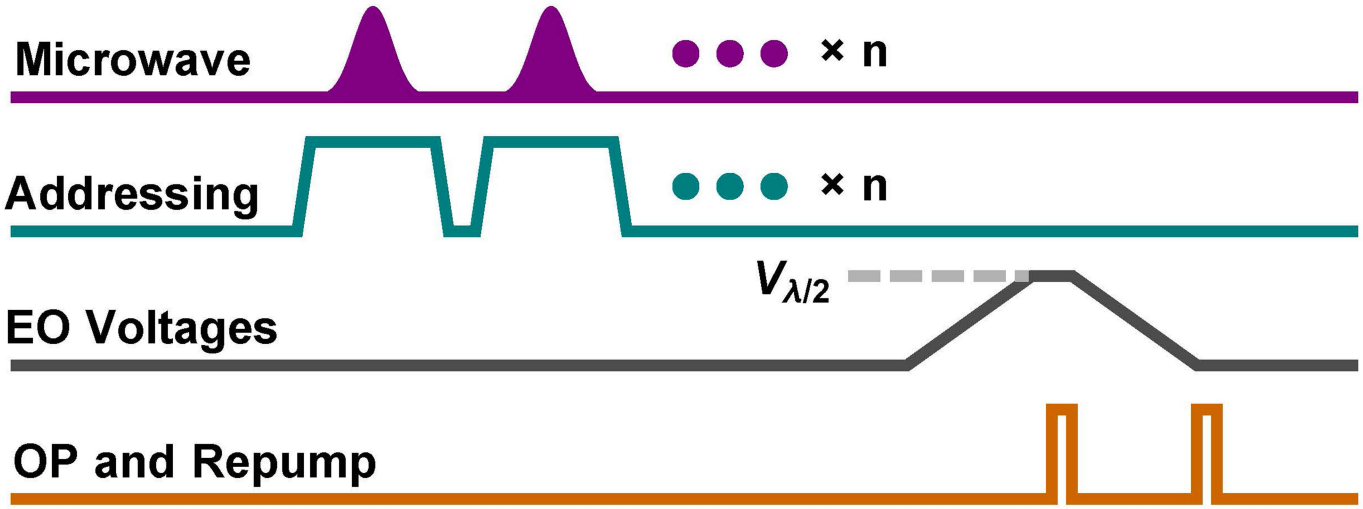
Monte Carlo simulations. Monte Carlo simulations of this sorting algorithm start with a randomly half-filled 3D array. Errors are probabilistically applied at each motion step and atom loss is considered after the completion of a sort. We calculate a separate motion fidelity for each internal state as the average of the measured fidelities in the three directions. One thousand simulations were run for various lattice dimensions and various target sublattices. For filling a $5\times5\times2$ or $4\times4\times3$ sublattice from a half-filled $5\times5\times5$ lattice, the simulations predict an average filling factor of about 0.97 after three sorts, in agreement with our measured filling factor to within the uncertainty associated with our measured errors.

Real-time control. The sorting process requires changing the timing sequence in real time. This is accomplished by combining real-time data analysis with two field-programmable gate arrays (FPGAs). The experiment has a 'backbone' of a fixed timing sequence. After the motion steps have been generated according to the initial occupancy map, the FPGAs pause that fixed timing sequence and take

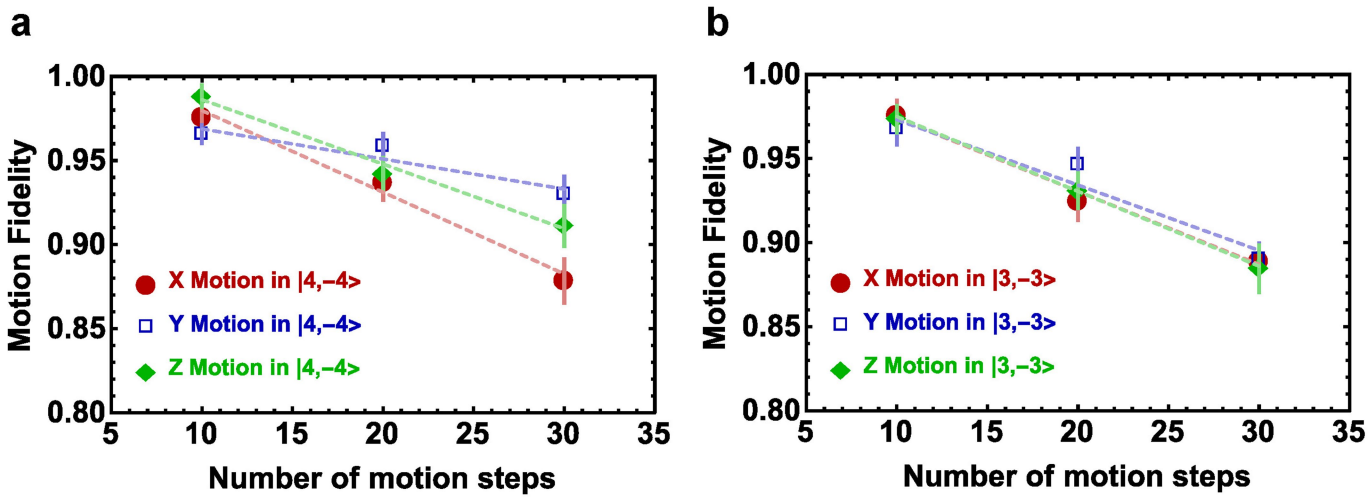
control of the electronic channels (optical pumping, electro-optic modulator voltages, addressing, microwaves) required for sorting. The data used for sorting, which comprise a sequence of directions for the motion steps and the lattice sites to be addressed at each motion step, are communicated to the FPGAs by the program that generates the occupancy map and creates the sorting plan. The FPGAs convert the motion steps into several voltage sequences that are output synchronously. After the motion steps have been executed, the FPGAs transfer the timing control back to the fixed backbone, which resumes where it was paused.

Code availability. The Monte Carlo code used to model our algorithm is available from the corresponding author on request.

Data availability. The underlying data used to generate the figures and conclusions in the paper are available from the corresponding author on reasonable request.



Extended Data Fig. 1 | Motion step. A motion step to move *n* atoms is shown. *n* atoms are sequentially targeted by the addressing beams and transferred from the 'stationary' state to the 'motion' state using microwaves. The electro-optic modulator (EO) voltages are ramped up to the half-wave voltage ($V_{\lambda/2}$) in order to move atoms by half of the lattice spacing. After motion, the atoms are optically pumped so that they all return to the stationary state. The EO voltages are then ramped back down. A final optical pumping (OP) ensures optimal preparation for the next motion step.



Extended Data Fig. 2 | Motion fidelities. a, **b**, Measured motion fidelity as a function of the number of motion steps in $|F = 4, m_F = -4\rangle$ (**a**) and $|F = 3, m_F = -3\rangle$ (**b**) in the x (maroon circles), y (blue squares) and z (green diamonds) directions. The lines are fits to the data. The error bars represent one standard deviation. Each point corresponds to about 600 atoms.

LETTER

https://doi.org/10.1038/s41586-018-0451-1

All-inorganic perovskite nanocrystal scintillators

Qiushui Chen[1], Jing Wu[2], Xiangyu Ou[3,4], Bolong Huang[5], Jawaher Almutlaq[6], Ayan A. Zhumekenov[6], Xinwei Guan[6], Sanyang Han[1], Liangliang Liang[1], Zhigao Yi[1], Juan Li[3,4], Xiaoji Xie[7], Yu Wang[8], Ying Li[8], Dianyuan Fan[8], Daniel B. L. Teh[9], Angelo H. All[9,10], Omar F. Mohammed[6], Osman M. Bakr[6], Tom Wu[11], Marco Bettinelli[12], Huanghao Yang[3,4]*, Wei Huang[7,13,14,15]* & Xiaogang Liu[1,8,16]*

The rising demand for radiation detection materials in many applications has led to extensive research on scintillators[1–3]. The ability of a scintillator to absorb high-energy (kiloelectronvolt-scale) X-ray photons and convert the absorbed energy into low-energy visible photons is critical for applications in radiation exposure monitoring, security inspection, X-ray astronomy and medical radiography[4,5]. However, conventional scintillators are generally synthesized by crystallization at a high temperature and their radioluminescence is difficult to tune across the visible spectrum. Here we describe experimental investigations of a series of all-inorganic perovskite nanocrystals comprising caesium and lead atoms and their response to X-ray irradiation. These nanocrystal scintillators exhibit strong X-ray absorption and intense radioluminescence at visible wavelengths. Unlike bulk inorganic scintillators, these perovskite nanomaterials are solution-processable at a relatively low temperature and can generate X-ray-induced emissions that are easily tunable across the visible spectrum by tailoring the anionic component of colloidal precursors during their synthesis. These features allow the fabrication of flexible and highly sensitive X-ray detectors with a detection limit of 13 nanograys per second, which is about 400 times lower than typical medical imaging doses. We show that these colour-tunable perovskite nanocrystal scintillators can provide a convenient visualization tool for X-ray radiography, as the associated image can be directly recorded by standard digital cameras. We also demonstrate their direct integration with commercial flat-panel imagers and their utility in examining electronic circuit boards under low-dose X-ray illumination.

The nature of the atomic constituents of a scintillator plays an important role in the radioluminescence process of the material because X-ray absorption increases exponentially with atomic number[6]. Although a wide range of scintillation materials containing heavy atoms have been characterized in detail for efficient X-ray scintillation, almost all of these materials are bulk crystals and grown by the Czochralski method[7] at temperatures above 1,700 °C. For bulk-form scintillators, such as $PbWO_4$ and $Bi_4Ge_3O_{12}$, a certain distance of exciton migration is typically needed to transport charge carriers for subsequent trapping by luminescence centres[8]. However, excessive exciton migration is detrimental because it can cause either radioluminescence afterglow or low-efficiency X-ray scintillation. In addition, conventional activator-doped scintillators, such as thallium-activated CsI (CsI:Tl) and cerium-activated $YAlO_3$ ($YAlO_3$:Ce), cannot produce tunable scintillation because of their fixed transition energies[9,10]. Despite enormous efforts, the development of scintillating materials that are low-temperature- and solution-processable, highly sensitive to X-rays and integrable to flexible substrates remains a daunting challenge.

Recently, bulk crystals of organic–inorganic hybrid perovskites have been found to exhibit large X-ray stopping power[11–14] and the ability of efficiently converting X-ray photons into charge carriers[15–18]. The direct photon-to-current conversion can be attributed to the heavy Pb atom[19] and large electron–hole diffusion lengths available in organic–inorganic hybrid perovskites[20–25]. We reason that caesium lead halide perovskite nanocrystals[26], which feature heavy constituent elements and tunable electronic bandgaps in the visible range, could be a promising candidate for high-efficiency X-ray scintillation. An appealing aspect of these perovskite nanocrystals is that their unique electronic structures render highly emissive triplet excited states[27] and anomalous fast emission rates[28]. By virtue of the effect of quantum confinement and increased overlap of electron and hole wavefunctions, the spatial distribution of luminescence centres and X-ray-generated excitons can be confined within the Bohr radius of the nanocrystals. Here we report experimental investigations of multicolour X-ray scintillation from a series of all-inorganic perovskite nanocrystals and demonstrate their use for ultrasensitive X-ray sensing and low-dose digital X-ray technology.

In a typical bulk scintillator material, incident X-ray photons can interact with heavy atoms (for example, Pb, Tl or Ce) to produce a large number of hot electrons through the photoelectric effect[8]. These charge carriers are quickly thermalized to form low-energy excitons, which can subsequently be transported to defect centres or activators for radiative luminescence (Extended Data Fig. 1a). We thus predict that high-energy (kiloelectronvolt-scale) X-ray photons can be converted to numerous low-energy visible photons via direct bandgap emissions in lead halide perovskite nanocrystals (Fig. 1a). To validate this hypothesis, we prepared a series of perovskite nanocrystals ($CsPbX_3$, with X = Cl, Br or I) by controlling the reaction of Cs-oleate with different PbX_2 precursors via a hot-injection solution method[29] (Extended Data Fig. 2). Transmission electron micrograph imaging reveals a cubic shape of the as-synthesized nanocrystals, with an average size of 9.6 nm (Fig. 1b). Remarkably, under X-ray beam excitation the perovskite quantum dots (QDs) yield narrow and colour-tunable emissions (Fig. 1c, Extended Data Fig. 3). This unique property allows multicolour, high-efficiency X-ray scintillation to be realized (Fig. 1d, e, Extended Data Table 1). By contrast, the radioluminescence spectrum of conventional bulk scintillators (CsI:Tl, $PbWO_4$, $YAlO_3$:Ce and $Bi_4Ge_3O_{12}$) is almost invariable and exhibits a wide emission peak with a large full-width at half-maximum (Extended Data Fig. 1b). This inherent limitation of conventional scintillators makes it difficult to achieve multicolour visualization of X-ray irradiation.

[1]Department of Chemistry, National University of Singapore, Singapore, Singapore. [2]School of Science, China University of Geosciences, Beijing, China. [3]MOE Key Laboratory for Analytical Science of Food Safety and Biology, Fuzhou University, Fuzhou, China. [4]State Key Laboratory of Photocatalysis on Energy and Environment, Fuzhou University, Fuzhou, China. [5]Department of Applied Biology and Chemical Technology, The Hong Kong Polytechnic University, Hong Kong, China. [6]Division of Physical Sciences and Engineering, King Abdullah University of Science and Technology, Thuwal, Saudi Arabia. [7]Institute of Advanced Materials, Nanjing Tech University, Nanjing, China. [8]SZU-NUS Collaborative Innovation Center for Optoelectronic Science and Technology, Shenzhen University, Shenzhen, China. [9]Singapore Institute for Neurotechnology, National University of Singapore, Singapore, Singapore. [10]Department of Biomedical Engineering, Johns Hopkins School of Medicine, Baltimore, MD, USA. [11]School of Materials Science and Engineering, University of New South Wales, Sydney, New South Wales, Australia. [12]Luminescent Materials Laboratory, DB, University of Verona, Verona, Italy. [13]Key Laboratory for Organic Electronics and Information Displays, Nanjing University of Posts and Telecommunications, Nanjing, China. [14]Institute of Advanced Materials, Nanjing University of Posts and Telecommunications, Nanjing, China. [15]Shaanxi Institute of Flexible Electronics, Northwestern Polytechnical University, Xi'an, China. [16]Center for Functional Materials, NUS Suzhou Research Institute, Suzhou, Jiangsu, China. *e-mail: hhyang@fzu.edu.cn; iamwhuang@nwpu.edu.cn; chmlx@nus.edu.sg

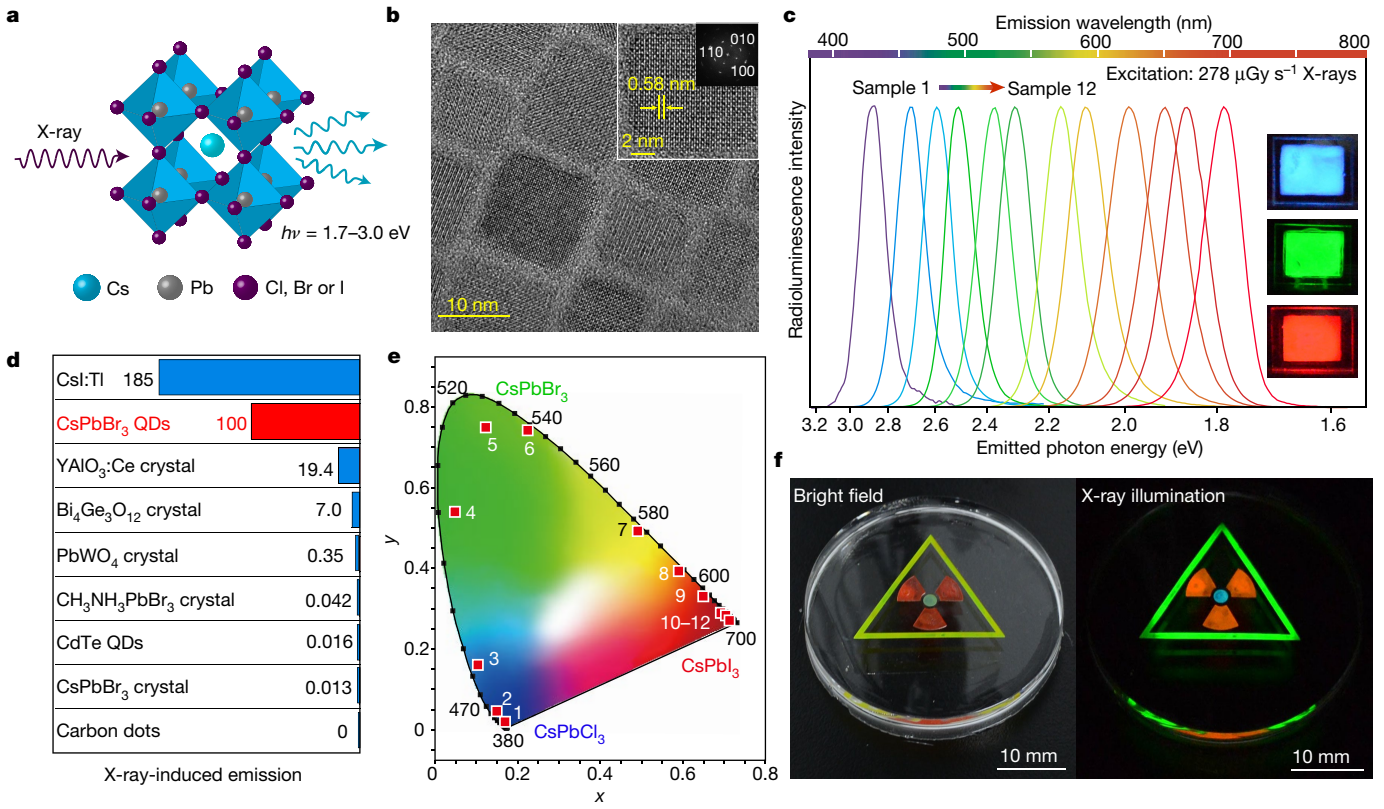


Fig. 1 | Full-colour radioluminescence from perovskite nanocrystal scintillators. a, Schematic representation of X-ray-induced luminescence of energy $h\nu$ (where h is the Planck constant and ν is the frequency), generated in an all-inorganic perovskite lattice with a cubic crystal structure (see main text for details). **b**, Low-resolution transmission electron microscopy (TEM) image of the as-synthesized $CsPbBr_3$ nanocrystals. The inset shows a high-resolution TEM image of a single $CsPbBr_3$ nanocrystal and the corresponding electron diffraction pattern along the [100] zoom axis. **c**, Tunable luminescence spectra of the perovskite QDs under X-ray illumination with a dose rate of 278 μGy s^{-1} at a voltage of 50 kV. The material compositions of samples 1–12 are $CsPbCl_3$ (1), $CsPbCl_2Br$ (2), $CsPbCl_{1.5}Br_{1.5}$ (3), $CsPbClBr_2$ (4), $CsPbCl_{2.5}Br_{0.5}$ (5), $CsPbBr_3$ (6), $CsPbBr_2I$ (7), $CsPbBr_{1.8}I_{1.2}$ (8), $CsPbBr_{1.5}I_{1.5}$ (9), $CsPbBr_{1.2}I_{1.8}$ (10), $CsPbBrI_2$ (11) and $CsPbI_3$ (12). The insets show photographs of the thin-film samples 3, 6 and 9, which emit blue, green and red colours, respectively, upon X-ray irradiation. **d**, Comparison of the optical sensitivity of various scintillator materials in response to exposure to X-rays produced at a voltage of 10 kV. **e**, CIE (Commission Internationale de l'Eclairage) chromaticity coordinates of the X-ray-induced visible emissions measured for samples 1–12. **f**, Multicolour X-ray scintillation (left, bright-field imaging; right, X-ray illumination at a voltage of 50 kV) from three types of perovskite nanocrystal scintillator (orange, $CsPbBr_2I$; green, $CsPbBr_3$; blue, $CsPbClBr_2$).

Inspired by the bandgap-tunable perovskite nanocrystal scintillators, we successfully developed a flexible prototype device for multicolour X-ray scintillation through a combination of solution processing and soft lithography (Fig. 1f, Extended Data Fig. 3d, e). The fabrication of the X-ray-sensing device was made possible by casting the oleate-capped perovskite nanocrystals onto the flexible substrate of interest. This flexible substrate allowed rapid X-ray multicolour visualization (Supplementary Video 1), which is inaccessible by current bulk scintillators.

We then compared the sensitivity of the perovskite nanocrystals to X-ray illumination with that of several of the most widely used commercial bulk scintillators (CsI:Tl, $PbWO_4$, $YAlO_3$:Ce and $Bi_4Ge_3O_{12}$). We used low-dose irradiation of 5.0 μGy s^{-1} (all doses refer to doses in air) at 10 kV and 5 μA and found that the ability of $CsPbBr_3$ nanocrystal thin films (thickness of about 0.1 mm) to convert X-ray photons into visible luminescence is comparable to that of high-efficiency CsI:Tl bulk scintillators (thickness of 5.0 mm), whereas it compares much more favourably (more intense by a factor of 5 or more) than other bulk scintillators, including $PbWO_4$, $YAlO_3$:Ce and $Bi_4Ge_3O_{12}$ (Fig. 1d). This superior performance is attributed to the large X-ray stopping power and high emission quantum yields of the lead halide QDs. Notably, conventional QDs (for example, CdTe QDs and carbon dots) exhibit low-efficiency X-ray-induced luminescence possibly due to weak X-ray absorption[30], and thus are not suitable for practical use as scintillators (Fig. 1d, Extended Data Fig. 4). As a point of comparison, we also found that typical bulk single crystals of $CsPbBr_3$ and $CH_3NH_3PbBr_3$ do not exhibit noticeable visible emission under the same experimental conditions (Fig. 1d, Extended Data Fig. 5). The noteworthy scintillation performance of $CsPbBr_3$ nanocrystals with respect to their bulk counterparts presents a compelling case for investigating the origins of the scintillation process in our system. This process can be explained in part by the lack of exciton confinement in the bulk form, in which discrete or quantized energy levels that give access to visible emission cannot be generated[31].

We further investigated experimentally and theoretically the physical processes that govern the interaction between X-rays and perovskite nanocrystals. As shown in Fig. 2a, we compared the absorption coefficient of the $CsPbBr_3$ nanocrystals (highest atomic number $Z_{max} = 82$; $K\alpha = 88.0$ keV) as a function of X-ray photon energy with two types of conventional QD (CdTe, $Z_{max} = 52$, $K\alpha = 31.8$ keV; carbon, $Z_{max} = 6$, $K\alpha = 0.285$ keV). The nature of heavy atomic constituents is critically important for efficient X-ray scintillation, because X-ray absorption scales with the effective atomic number, Z_{eff}, as Z_{eff}^4/AE^3, where A is the atomic mass and E is the X-ray photon energy[6]. We thus speculate that the Pb-based perovskite nanocrystals are much more suitable for efficient X-ray absorption than QDs without the Pb component. We carried out an X-ray photoelectron spectroscopic investigation to record the kinetic process of electrons escaping from the $CsPbBr_3$ nanocrystal upon irradiation with soft X-rays (Fig. 2b). To reveal the photoionization nature of the X-ray scintillation process under study, we

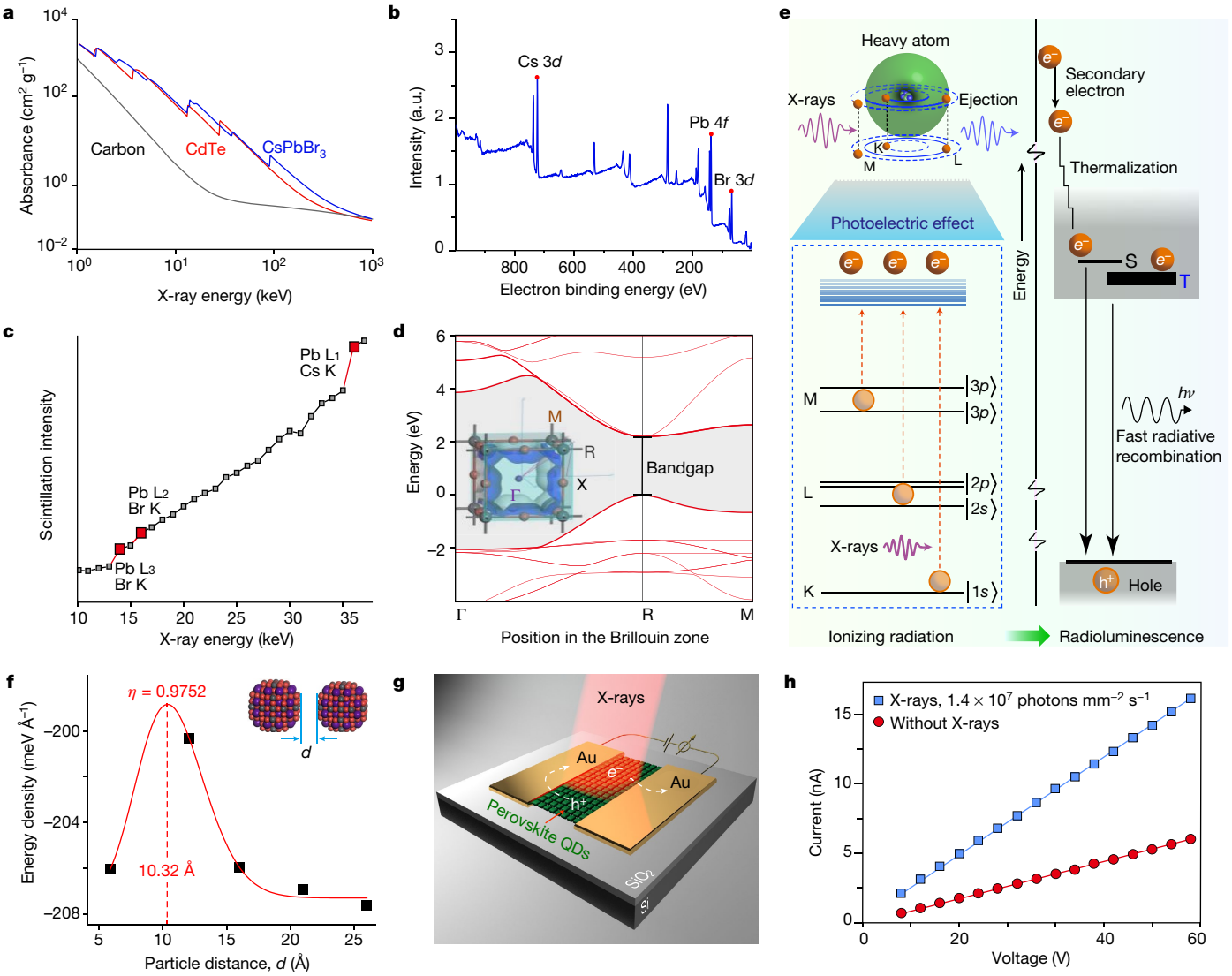


Fig. 2 | Mechanistic investigation of X-ray energy conversion by perovskite nanocrystals. a, Measured absorption spectra of $CsPbBr_3$, CdTe and carbon as a function of X-ray energy. The attenuation coefficients were obtained from ref. [33]. **b**, X-ray photoelectron spectroscopic data of the $CsPbBr_3$ nanocrystals plotted against the binding energy of the electron. The photoemission peaks Cs 3*d*, Pb 4*f* and Br 3*d* are indicated. a.u., arbitrary units. **c**, Measurement of X-ray-induced luminescence from the perovskite nanocrystals using synchrotron radiation. The electronic edge energies of Pb L, Cs K and Br K (shown as red squares) fall in the X-ray energy range 10–38 keV. The line is a guide for the eye. **d**, Calculated electronic band structures of the $CsPbBr_3$ nanocrystal. The inset shows the Brillouin zone of the cubic-phased crystal lattice (see Methods for details). **e**, Proposed mechanism of X-ray scintillation in a lead halide perovskite nanocrystal. Upon X-ray irradiation, a high-energy electron (red circles, e^-) is ejected from a lattice atom through photoelectric ionization (ionizing radiation creates an energetic electron and a hole in an inner electronic shell). Subsequently, the ejected high-energy electron produces secondary high-energy electrons. The generated hot charge carriers then undergo thermalization and produce low-energy excitons. Next, fast radiative recombination takes place, producing radioluminescence of energy $h\nu$ in either a singlet (S) or triplet (T) state at the electronic band edge. **f**, Energy density on the surface of a $CsPbBr_3$ cluster as a function of particle distance, *d*, in the lattice. The red line is a fit with a Gaussian distribution function with a fitting coefficient of $\eta = 0.9752$. The particle distance corresponding to the maximum energy density is 10.32 Å. **g**, Schematic showing the basic design of a perovskite-nanocrystal-based photoconductor used for X-ray sensing. A 10-μm-thick layer of $CsPbBr_3$ QDs is spin-coated onto the substrate for X-ray photon–carrier conversion. Gold (Au) electrodes are placed onto the QDs for hole–electron extraction. **h**, Current–voltage characteristics of the as-fabricated photoconductor, recorded with and without X-ray illumination.

measured the radioluminescence of the perovskite nanocrystals in response to synchrotron radiation (Fig. 2c, Extended Data Fig. 6, Supplementary Video 2). We observed an abrupt enhancement in the scintillation intensity upon excitation at 14 keV, 16 keV and 36 keV, indicating an X-ray absorption resonance at the electronic edge of the Pb L, Cs K and Br K shells in the $CsPbBr_3$ structure. Density functional theory calculations confirmed that the electronic band structure of these perovskite nanocrystal scintillators is tunable, which is associated with the tailorability of their valence band through control of the halide composition (Extended Data Fig. 7). The bandgap energy of the perovskite nanocrystal under study is located in the range 1.7–3 eV, suggesting the feasibility of using such a nanomaterial to convert an absorbed dose of ionizing radiation into visible light (Fig. 2d). In addition, the orbital contour plots of the $CsPbBr_3$ nanocrystal indicate that the presence of hole-like surface-vacancy-induced Coulomb-trapping states near the Fermi level beyond the valence band maximum is responsible for the electronically energetic confinement of excitons in the perovskite nanocrystal (Extended Data Fig. 8b).

Figure 2e presents a plausible mechanism for the high-intensity radioluminescence from the perovskite nanocrystals. At the initial conversion stage, an incident X-ray photon with energy lower than a few hundred kiloelectronvolts interacts with the lattice atoms of a perovskite nanocrystal, predominantly through the photoelectric effect. During this process a large number of high-energy electrons and holes

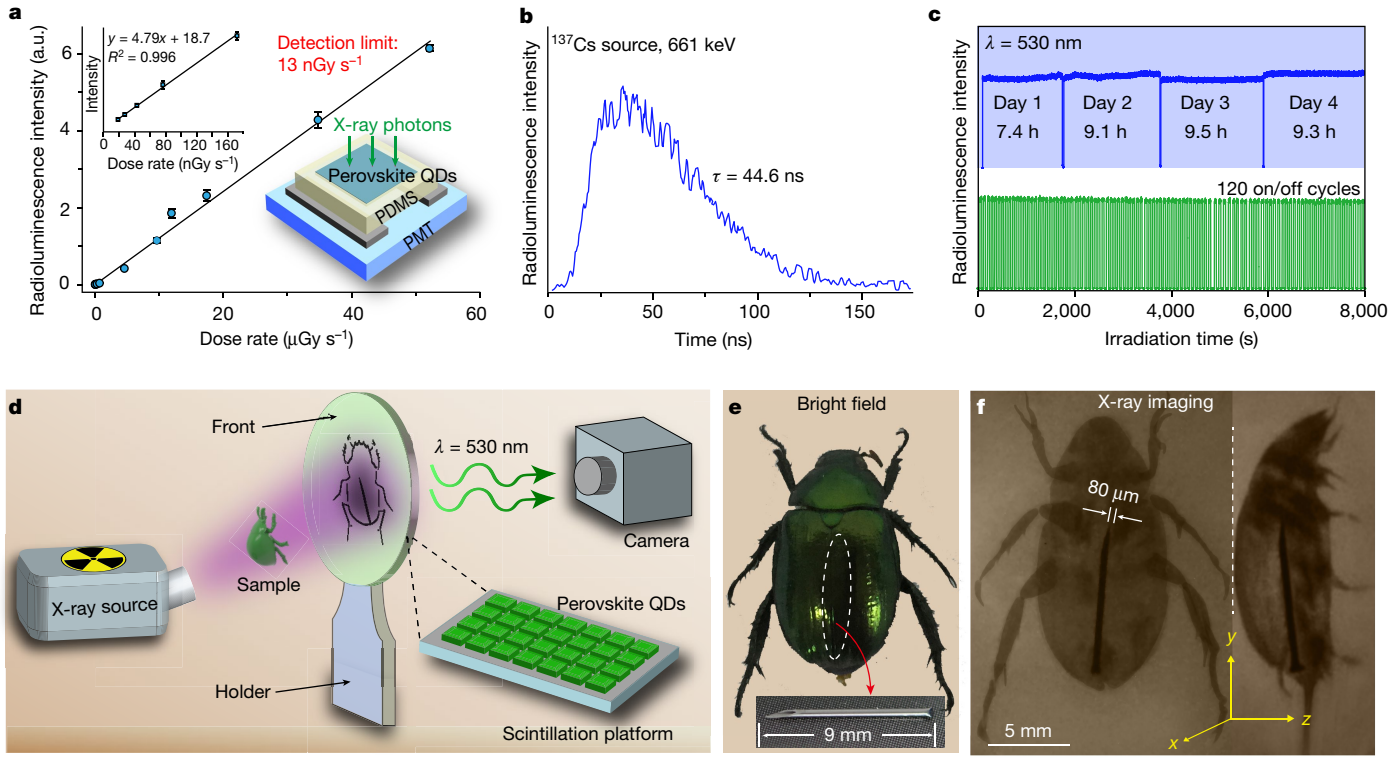


Fig. 3 | Ultrasensitive X-ray sensing and radiography using $CsPbBr_3$ nanocrystals. a, Radioluminescence measurements for a $CsPbBr_3$-based scintillator as a function of dose rate. The left inset shows radioluminescence profiles measured at low dose rates. The detection limit of 13 nGy s^{-1} is derived from the slope of the fitting line, with a signal-to-noise ratio of 3. The right inset shows a schematic of the X-ray photodetector, which consists of a $CsPbBr_3$ nanocrystal thin film (about 120 μm thickness), a polydimethylsiloxane (PDMS) layer and a photomultiplier tube (PMT). All measurements were performed three times. Error bars are mean ± s.d. **b**, Measured radioluminescence decay of the $CsPbBr_3$-based scintillator under excitation with a ^{137}Cs source (photon energy, 661 keV). The scintillation decay time is $\tau = 44.6$ ns. **c**, Photostability of the $CsPbBr_3$-based scintillator against continuous X-ray irradiation (wavelength $\lambda = 530$ nm, 50 kV; top) and repeated cycles of X-ray excitation at 30 kV with a time interval of 30 s ($\lambda = 530$ nm; bottom). **d**, Schematic of the experimental setup used for real-time X-ray diagnostic imaging of biological samples. A beetle is placed between the X-ray source and a scintillation platform covered with perovskite QDs. **e**, **f**, Bright-field (**e**) and the X-ray (**f**) images of the sample, recorded with a digital camera. The X-ray images were recorded at a voltage of 50 kV.

can be created, and electronic transport occurs between the perovskite nanocrystals (Fig. 2f). The hot electrons and holes are then quickly thermalized in the conduction and valence band edges. The X-ray-induced charge carriers in the perovskite nanocrystals were experimentally confirmed by measuring the current through a photoconductor upon X-ray illumination (Fig. 2g, h). The trapping and radiative recombination of electron–hole pairs can be controlled to produce a desired luminescence colour by adjusting the bandgap energy. The mechanism of intense X-ray scintillation could be attributed in part to the strong X-ray stopping power and quantum confinement effects of perovskite nanocrystals. Additionally, the scintillation process is dominated by the presence of highly emissive triplet excited states (Fig. 2e), large absorption cross-section within the bandgap (Extended Data Fig. 9a, b) and fast emission output (Extended Data Fig. 9c–e), which are characteristics of perovskite nanocrystals[27,28,32].

The solution-processability of the perovskite nanocrystals makes it possible to fabricate a thin-film scintillator device for ultrasensitive X-ray detection. In this device (Fig. 3a), spin-coated $CsPbBr_3$ nanocrystals are used for X-ray sensing by converting high-energy X-ray photons into visible emission, which is readily detectable by a photomultiplier tube. A favourable characteristic of the prototype X-ray detector is its linear response to the X-ray dose rate, covering a range as broad as four orders of magnitude (Extended Data Fig. 10). The lowest detectable dose rate for X-ray detection is demonstrated to be 13 nGy s^{-1}. This value is about 420 times lower than the dose typically used for X-ray diagnostics (5.5 μGy s^{-1})[14]. This scintillation photodetector also exhibits a very fast response (scintillation decay time, $\tau = 44.6$ ns) upon excitation with pulsed photons (661 keV) from a portable ^{137}Cs source (Fig. 3b). The fast response to X-ray photons is critical to scintillation performance in medical radiography. The photostability of the perovskite nanocrystals was further examined under continuous or repeated cycles of X-ray illumination, as shown in Fig. 3c.

To assess the suitability of the perovskite nanocrystals as scintillators for X-ray phase-contrast imaging, we implanted a metallic needle into a green scarab beetle and imaged the biological specimen with X-rays against a background substrate comprising a thin film of solution-processed $CsPbBr_3$ nanocrystals (Fig. 3d). We note that the $CsPbBr_3$ nanocrystals were chosen for this demonstration because their green emission at 530 nm matches well with the maximum wavelength response of a complementary metal-oxide-semiconductor sensor. As shown in Fig. 3e, f, owing to the large difference between the X-ray stopping powers of the needle and the beetle, the needle inside the beetle is clearly revealed by phase-contrast imaging recorded using a common digital camera. The concept of direct X-ray contrast imaging through the use of high-efficiency perovskite nanocrystals is readily applicable to high-throughput electronics inspection and tissue imaging, where common digital cameras can be conveniently used (Extended Data Fig. 11; Extended Data Table 2).

We took a step further and tested the compliance of the perovskite nanocrystals to commercial flat-panel X-ray detectors equipped with α-Si photodiode arrays (Fig. 4a, b). As shown in Fig. 4c, the perovskite-nanocrystal-based X-ray detector shows a modulation transfer function of 0.72 at a spatial resolution of 2.0 line pairs per millimetre, which is much higher than the spatial resolution of commercially used CsI:Tl-based flat-panel X-ray detectors (0.36 at 2.0 line pairs per millimetre). This high spatial resolution could be ascribed to the lower degree of light scattering in the nanoparticle-based thin film

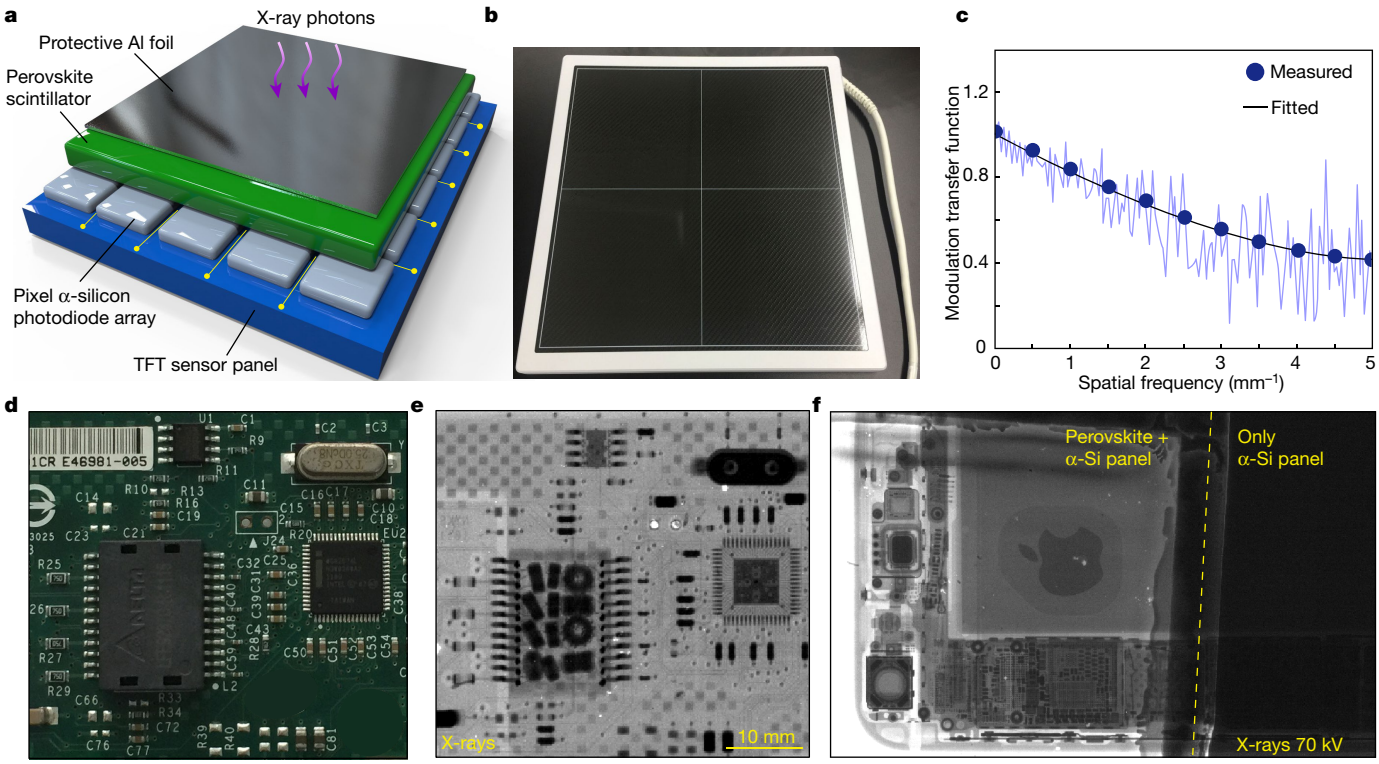


Fig. 4 | Prototype perovskite-nanocrystal-based flat-panel detector for digital radiography. a, Multilayered design of the flat-panel X-ray imaging system consisting of a thin-film-transistor (TFT) sensor panel, a pixelated α-silicon photodiode array, a $CsPbBr_3$ perovskite nanocrystal thin film (about 75 μm thick) and a protective Al foil cover (40 μm). **b**, Photograph of the packaged flat-panel detector. **c**, Spatial resolution of the X-ray imaging system, characterized by the modulation transfer function under 15 μGy of X-ray exposure. The blue circles and purple line show measured values and the black line is a fit to the data. **d**, **e**, Digital photograph of a network interface card (**d**) and corresponding X-ray image obtained using the flat-panel detector (70 kV and 2.5 mGy s^{-1} exposure for 6 ms) (**e**). **f**, Comparison of X-ray images of an Apple iPhone acquired with the perovskite scintillator deposited on an α-Si photodiode panel (left) and only with an α-Si photodiode (right).

compared with that occurring in commercial bulk-scintillator-based films made of thick polycrystalline ceramics or long micropillars. We further used the prototype device to image the internal structures of electronic circuits and an Apple iPhone with a low X-ray dose of 15 μGy (Fig. 4d–f). Unlike CsI:Tl scintillators, which have the issue of afterglow luminescence (scintillation decay time of 1,000 ns), our perovskite nanocrystals have a very fast response (44.6 ns) to X-rays, making them ideal for dynamic real-time X-ray imaging.

In conclusion, we have demonstrated inorganic perovskite nanocrystals as a new class of scintillators that are capable of converting small doses of X-ray photons into multicolour visible light. When considering the material's solution-processability and practical scalability, it is envisioned that these scintillators are suitable for the mass production of ultrasensitive X-ray detectors and large-area, flexible X-ray imagers. Compared to conventional CsI:Tl scintillators—whose use is constrained by the risk of thallium poisoning, the presence of afterglow and high-temperature synthesis—perovskite nanocrystals offer several outstanding attributes, including relatively low toxicity, low-temperature solution synthesis, fast scintillation response and high emission quantum yield. Although there is still much to be learned regarding the origin of nanocrystal scintillation, these perovskite nanocrystals may hold substantial promise for advancing X-ray sensing and imaging industry. The thermal and environmental instability issues that are often associated with perovskite materials in photovoltaic and light-emitting-diode applications could be largely avoided through the X-ray scintillation settings.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, statements of data availability and associated accession codes are available at https://doi.org/10.1038/s41586-018-0451-1.

Received: 24 October 2017; Accepted: 21 June 2018;
Published online 27 August 2018.

1. Röntgen, W. C. On a new kind of rays. *Science* **3**, 227–231 (1896).
2. Moretti, F. et al. Radioluminescence sensitization in scintillators and phosphors: trap engineering and modeling. *J. Phys. Chem. C* **118**, 9670–9676 (2014).
3. Büchele, P. et al. X-ray imaging with scintillator-sensitized hybrid organic photodetectors. *Nat. Photon.* **9**, 843–848 (2015).
4. Yaffe, M. J. & Rowlands, J. A. X-ray detectors for digital radiography. *Phys. Med. Biol.* **42**, 1–39 (1997).
5. Durie, B. G. & Salmon, S. E. Scintillator distribution in high-speed autoradiography. *Science* **190**, 1093–1095 (1975).
6. Nikl, M. & Yoshikawa, A. Recent R&D trends in inorganic single-crystal scintillator materials for radiation detection. *Adv. Opt. Mater.* **3**, 463–481 (2015).
7. Weber, M. J. Inorganic scintillators: today and tomorrow. *J. Lumin.* **100**, 35–45 (2002).
8. Rodnyi, P. A. *Physical Processes in Inorganic Scintillators* (CRC Press, Boca Raton, 1997).
9. Nagarkar, V. V. et al. Structured CsI(Tl) scintillators for X-ray imaging applications. *IEEE Trans. Nucl. Sci.* **45**, 492–496 (1998).
10. Baccaro, S. et al. Scintillation properties of YAP:Ce. *Nucl. Instrum. Methods A* **361**, 209–215 (1995).
11. Rowlands, J. A. Material change for X-ray detectors. *Nature* **550**, 47–48 (2017).
12. Kim, Y. C. et al. Printable organometallic perovskite enables large-area, low-dose X-ray imaging. *Nature* **550**, 87–91 (2017).
13. Wei, W. et al. Monolithic integration of hybrid perovskite single crystals with heterogenous substrate for highly sensitive X-ray imaging. *Nat. Photon.* **11**, 315–321 (2017).
14. Wei, H. et al. Sensitive X-ray detectors made of methylammonium lead tribromide perovskite single crystals. *Nat. Photon.* **10**, 333–339 (2016).
15. Yakunin, S. et al. Detection of X-ray photons by solution-processed organic-inorganic perovskites. *Nat. Photon.* **9**, 444–449 (2015).
16. Shrestha, S. et al. High-performance direct conversion X-ray detectors based on sintered hybrid lead triiodide perovskite wafers. *Nat. Photon.* **11**, 436–440 (2017).
17. Wei, H. et al. Dopant compensation in alloyed $CH_3NH_3PbBr_{3-x}Cl_x$ perovskite single crystals for gamma-ray spectroscopy. *Nat. Mater.* **16**, 826–833 (2017).
18. Pan, W. et al. $Cs_2AgBiBr_6$ single-crystal X-ray detectors with a low detection limit. *Nat. Photon.* **11**, 726–732 (2017).

19. Birowosuto, M. D. et al. X-ray scintillation in lead halide perovskite crystals. *Sci. Rep.* **6**, 37254 (2016).
20. Dong, Q. et al. Electron–hole diffusion lengths >175 μm in solution-grown $CH_3NH_3PbI_3$ single crystals. *Science* **347**, 967–970 (2015).
21. Shi, D. et al. Low trap-state density and long carrier diffusion in organolead trihalide perovskite single crystals. *Science* **347**, 519–522 (2015).
22. Tan, H. et al. Efficient and stable solution-processed planar perovskite solar cells via contact passivation. *Science* **355**, 722–726 (2017).
23. Li, X. et al. A vacuum flash-assisted solution process for high-efficiency large-area perovskite solar cells. *Science* **353**, 58–62 (2016).
24. Im, J.-H., Jang, I.-H., Pellet, N., Grätzel, M. & Park, N.-G. Growth of $CH_3NH_3PbI_3$ cuboids with controlled size for high-efficiency perovskite solar cells. *Nat. Nanotechnol.* **9**, 927–932 (2014).
25. Son, D.-Y. et al. Self-formed grain boundary healing layer for highly efficient $CH_3NH_3PbI_3$ perovskite solar cells. *Nat. Energy* **1**, 16081 (2016).
26. Protesescu, L. et al. Nanocrystals of cesium lead halide perovskites ($CsPbX_3$, X=Cl, Br, and I): novel optoelectronic materials showing bright emission with wide color gamut. *Nano Lett.* **15**, 3692–3696 (2015).
27. Becker, M. A. et al. Bright triplet excitons in caesium lead halide perovskites. *Nature* **553**, 189–193 (2018).
28. Hu, F. et al. Superior optical properties of perovskite nanocrystals as single photon emitters. *ACS Nano* **9**, 12410–12416 (2015).
29. Swarnkar, A. et al. Quantum dot-induced phase stabilization of α-$CsPbI_3$ perovskite for high-efficiency photovoltaics. *Science* **354**, 92–95 (2016).
30. Hossu, M., Liu, Z., Yao, M., Ma, L. & Chen, W. X-ray luminescence of CdTe quantum dots in LaF_3:Ce/CdTe nanocomposites. *Appl. Phys. Lett.* **100**, 013109 (2012).
31. Saidaminov, M. I. et al. Pure Cs_4PbBr_6: highly luminescent zero-dimensional perovskite solids. *ACS Energy Lett.* **1**, 840–845 (2016).
32. Kovalenko, M. V., Protesescu, L. & Bodnarchuk, M. I. Properties and potential optoelectronic applications of lead halide perovskite nanocrystals. *Science* **358**, 745–750 (2017).
33. Berger, M. J. et al. *XCOM: Photon Cross Sections Database*; https://www.nist.gov/pml/xcom-photon-cross-sections-database (2013).

Acknowledgements This work is supported by the King Abdullah University of Science and Technology; the Singapore Ministry of Education (grants R143000627112 and R143000642112); the Agency for Science, Technology and Research (A*STAR) under contracts 122-PSE-0014 and 1231AFG028 (Singapore); the National Research Foundation, Prime Minister's Office, Singapore under its Competitive Research Program (CRP award number NRF-CRP15-2015-03); the National Basic Research Program of China (973 Program, grant number 2015CB932200); the National Natural Science Foundation of China (21635002, 21471109, 21210001 and 21405143); and the Natural Science Foundation of Jiangsu Province (BE2015699). We thank H. Jiang, B. Deng, Z. Fang, Z. Zhou, Y. Zhang, X. Ling, M. Sun and A. Malko for technical assistance.

Reviewer information *Nature* thanks R. Comin, W. Heiss and the other anonymous reviewer(s) for their contribution to the peer review of this work.

Author contributions Q.C. and X.L. conceived and initiated the project. X.L., H.Y. and W.H. supervised the project and led the collaboration efforts. Q.C., X.L., H.Y. and W.H. designed the experiments. Q.C., J.W., L.L. and S.H. performed the nanocrystal synthesis. Q.C., X.O. and J.L. carried out the spectral measurements. Q.C., X.O., Y.W., Y.L., D.F., Z.Y., D.B.L.T. and A.H.L. contributed to the design and implementation of the X-ray sensing experiments. B.H., M.B. and O.F.M. carried out the theoretical calculations. J.A., A.A.Z. and O.M.B. prepared the perovskite single crystals. X.G. and T.W. fabricated the photoconductor devices and performed the photocurrent measurements. X.X. fabricated the PDMS moulds and measured the low-temperature scintillation spectra. Q.C. and X.L. wrote the manuscript. All authors discussed the results and commented on the manuscript.

Competing interests The authors declare no competing interests.

Additional information
Extended data is available for this paper at https://doi.org/10.1038/s41586-018-0451-1.
Supplementary information is available for this paper at https://doi.org/10.1038/s41586-018-0451-1.
Reprints and permissions information is available at http://www.nature.com/reprints.
Correspondence and requests for materials should be addressed to H.Y. or W.H. or X.L.
Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

METHODS

Chemicals. Caesium carbonate (Cs_2CO_3, 99.9%), lead(II) chloride ($PbCl_2$, 99.99%), lead(II) bromide ($PbBr_2$, 99.99%), lead(II) iodide (PbI_2, 99.99%), oleylamine (technical grade 70%), oleic acid (technical grade 90%), 1-octadecene (technical grade 90%) and cyclohexane (chromatography grade 99.9%) were purchased from Sigma-Aldrich. Silicon wafers were obtained from Xilika Crystal Polishing Material Co., Ltd (Tianjin, China). SU-8 photoresist (2050) and developer solution were purchased from Microchem Corp. (Newton, MA). A Sylgard 184 silicone elastomer kit was purchased from Dow Corning for the preparation of polydimethylsiloxane (PDMS) substrates. Crystals of CsI:Tl, $Bi_4Ge_3O_{12}$, $YAlO_3$:Ce and $PbWO_4$ scintillators were purchased from Zhonghelixin Co., Ltd (Chengdu, China). CdTe QDs were obtained from Xingzi New Material Technology Development Co., Ltd (Shanghai, China). Unless otherwise noted, all the reagents were used without additional treatment.

Synthesis of Cs-oleate as a caesium precursor. In a typical synthesis procedure, Cs_2CO_3 (0.4 g, 1.23 mmol), oleic acid (1.25 ml) and octadecene (15 ml) were added to a two-neck round-bottom flask (50 ml). The resulting mixture was heated to 100 °C under vigorous stirring and vacuum conditions for 0.5 h. After that, a nitrogen purge and vacuum were alternately applied to the flask three times to remove moisture and O_2. Subsequently, the mixture was heated to 150 °C and the solution became clear, indicating the completion of the reaction between Cs_2CO_3 and oleic acid. The Cs-precursor solution was kept at 150 °C in a nitrogen atmosphere before the synthesis of perovskite nanocrystal.

Synthesis of $CsPbX_3$ (X = Cl, Br or I) nanocrystals. The $CsPbX_3$ perovskite QDs were synthesized by a modified hot-injection procedure[29]. In a typical experiment, PbX_2 (0.36 mmol each for X = Cl, Br or I), oleic acid (1.0 ml), oleylamine (1.0 ml) and octadecene (10 ml) were added to a two-neck round-bottom flask (50 ml). The resulting mixture was heated to 100 °C under vigorous stirring and vacuum conditions for 0.5 h, at which time the moisture residue was removed by purging with nitrogen and vacuum suction. Then the mixture was heated to 160 °C until the PbX_2 precursors dissolved completely. A hot Cs-oleate precursor solution (1 ml) was injected quickly into the above reaction mixture. After 5 s of reaction, the flask was transferred into an ice bath. The $CsPbX_3$ QDs were obtained by centrifugation at 13,000 r.p.m. for 10 min and stored in 4 ml of cyclohexane before further use. Mixed-halide perovskite QDs were synthesized to tune the luminescence colour. Samples 1–12 correspond to the as-synthesized $CsPbCl_3$ (1), $CsPbCl_2Br$ (2), $CsPbCl_{1.5}Br_{1.5}$ (3), $CsPbClBr_2$ (4), $CsPbCl_{2.5}Br_{0.5}$ (5), $CsPbBr_3$ (6), $CsPbBr_2I$ (7), $CsPbBr_{1.8}I_{1.2}$ (8), $CsPbBr_{1.5}I_{1.5}$ (9), $CsPbBr_{1.2}I_{1.8}$ (10), $CsPbBrI_2$ (11) and $CsPbI_3$ (12) QDs.

Growth of lead halide perovskite single crystals. The growth of $CsPbBr_3$ single crystals was carried out according to a method described in the literature[34]. In a typical procedure, CsBr (0.64 g, 3 mmol) and $PbBr_2$ (2.2 g, 6 mmol) were dissolved in 3 ml of dimethyl sulfoxide and stirred for 1 h. Subsequently, 1.5 ml of the mixture was transferred into a vial heated at 60 °C, and the temperature was raised to 100 °C with a heating rate of 10 °C h^{-1}. At 100 °C, the solution was filtered and then gradually heated to 120 °C. We observed the growth of small-sized crystals with increasing temperature. The resulting crystals were washed with hot dimethyl sulfoxide and dried in vacuo at 100 °C for 1 h. For the synthesis of $CH_3NH_3PbBr_3$ ($MAPbBr_3$) single crystals, a mixture of $PbBr_2$ and CH_3NH_3Br (1.5 mol each) was dissolved in a solution of *N*,*N*,-dimethyl formamide (1 ml) at room temperature. The solution was purified by passing through a polytetrafluoroethylene filter with a pore size of 0.22 μm. The growth of the $MAPbBr_3$ single crystals was carried out in an oil bath heated at 60 °C and under ambient pressure.

Synthesis of fluorescent carbon dots. Fluorescent carbon dots were synthesized by a hydrothermal method. In a typical experiment, ammonium citrate (0.972 g, 4.0 mmol) was first dissolved in a 20-ml water solution. The solution was then transferred into a 30-ml Teflon-lined vessel at room temperature while stirring. Subsequently, the solution was heated to 190 °C and kept at that temperature for 10 h. After cooling to room temperature, the product was purified using dialysis, and the cut-off molecular weight of the dialysed membrane was equivalent to about 2,000. The carbon dot solution was concentrated using a rotary evaporator.

Preparation of silica-coated perovskite nanocrystals. The stability of perovskite nanocrystals was improved by coating with a silicon dioxide (silica) layer according to the literature[35]. In a typical procedure, the $CsPbBr_3$ QDs, dispersed in cyclohexane (2 ml), were introduced into a 50-ml flask containing 10 ml of toluene solution (≥99.5%; AnalaR NORMAPUR). 100 μl of tetramethoxysilane was injected quickly into the mixture at room temperature. After stirring for 2 h, the products were isolated through centrifugation at 13,000 r.p.m. for 8 min. The silica-coated $CsPbX_3$ (X = Br or I) QDs were either dispersed in cyclohexane or dried in air.

Physical characterization. Powder X-ray diffraction characterization was carried out by an ADDS wide-angle X-ray powder diffractometer with Cu Kα radiation (wavelength, λ = 1.54184 Å). TEM imaging was performed using a FEI Tecnai G20 transmission electron microscope with an accelerating voltage of 200 kV. X-ray photoelectron spectroscopy analysis was carried out using a Thermo escalab 250Xi instrument equipped with Al Kα monochromatized X-rays at 1,486.6 eV. Absorption spectra were measured by an ultraviolet–visible spectrophotometer (UV-2450, Shimadzu, Japan). Photoluminescence and radioluminescence spectra were obtained by an Edinburgh FS5 fluorescence spectrophotometer (Edinburgh Instruments Ltd, UK) equipped with a miniature X-ray source (AMPEK, Inc.). Photographs of the X-ray-induced luminescence were acquired with a digital camera (Nikon D7100 with AF Micro-Nikkor 60mm f/2.8D). For the time-resolved photoluminescence measurements, a pulsed excitation source was used. The scintillation decay measurement was carried out at the Institute of High Energy Physics of the Chinese Academy of Sciences with a ^{137}Cs source used for the pulsed excitation. The effective scintillation decay time (τ_{eff}) can be calculated using the following formula:

$$\tau_{eff} = \frac{1}{I_0}\int_0^{\infty} I(t)\mathrm{d}t$$

where $I(t)$ and I_0 denote the radioluminescence (or photoluminescence) intensity as a function of time, t, and the maximum intensity, respectively.

Measurement of photoluminescence quantum yield. The quantum yield was determined with an optical spectrometer equipped with an integrating sphere. Perovskite QDs were dispersed in cyclohexane. The excitation and luminescence emission were detected by a photomultiplier tube (PMT) through total internal reflection in the integration sphere. The photoluminescence quantum yield (PLQY) was calculated according to $PLQY = P_{sample}/(S_{ref} - S_{sample})$, where S_{ref} and S_{sample} are the excitation light intensities not absorbed by the solvent and the sample, respectively, and P_{sample} is the integrated emission intensity of the sample (Extended Data Fig. 5e).

Measurement of exciton binding energy in perovskite nanocrystal scintillator. The exciton binding energy (E_a) was estimated by measuring the temperature-dependent radioluminescence intensity. By fitting data derived from the integrated luminescence intensity of the $CsPbBr_3$ QD scintillator with the Arrhenius formula:

$$I(T) = \frac{I(T_0)}{1 + CT\exp[-E_a/(k_BT)]}$$

where $I(T_0)$ is the radioluminescence intensity at the low-temperature (T_0) limit, k_B is the Boltzmann constant, C is a constant and T is the temperature, we obtain an exciton binding energy of 49 meV.

Fabrication of perovskite nanocrystal scintillator films on PDMS substrates. The PDMS substrates were fabricated by a standard soft lithography microfabrication technique. Briefly, a photomask was first designed using Adobe Illustrator CS6. A 60-μm-thick layer of negative photoresists (SU-8 2015; 2,500 r.p.m., 60 s) was spin-coated onto a silicon wafer (3 inch; 1 inch = 1.54 cm). The wafer was prebaked at 60 °C for 10 min and then at 85 °C for 5 min. The resulting photoresist on the wafer was irradiated by an ultraviolet lamp for 20 s, followed by a post-baking treatment in an oven at 75 °C for 5 min. Next, the desired microstructure on the silicon wafer was produced using a developer solution. The PDMS substrates were fabricated with a premixed PDMS prepolymer and curing agent (10:1 by mass) under vacuum conditions, followed by heat treatment at 80 °C for 2 h. The PDMS replicas were carefully peeled off from the master. Finally, perovskite QDs dispersed in cyclohexane were coated onto the PDMS substrate.

Radioluminescence measurement for perovskite nanocrystal scintillators. The measurement of X-ray-induced luminescence was performed using a solid film comprising perovskite QDs. We note that perovskite QDs dispersed in solution are not suitable for scintillation characterization under X-ray excitation, because a low population of QDs in solution is inefficient for X-ray absorption. Unlike under visible-light excitation, a quartz cuvette is not used for measuring scintillation luminescence under X-ray excitation, because the excitation can be strongly absorbed by the cuvette. The scintillation decay times[36] of CsI:Tl, $Bi_4Ge_3O_{12}$, $YAlO_3$:Ce and $PbWO_4$ crystal scintillators are listed in Extended Data Table 1.

X-ray photoconductor devices. To fabricate the X-ray photoconducting device, silica wafers with a 300-nm-thick SiO_2 layer were first cleaned by sonication in acetone, ethanol and deionized water separately. After drying with flowing nitrogen, the substrates were treated with oxygen plasma for 6 min. The solution of $CsPbBr_3$ QDs was spin-coated onto the Si/SiO_2 substrates at 500 r.p.m. for 30 s and subsequently annealed at 100 °C for 5 min. This procedure was repeated three times to produce a film with a thickness of about 10 μm. After that, 100-nm-thick gold electrodes were deposited onto the $CsPbBr_3$ QD film by thermal evaporation, using a shadow mask to control the size of the deposition. For the X-ray photon-to-current measurement, we used a commercially available, miniaturized X-ray tube (Amptek). The target in the X-ray tube was made of gold and the maximum output was 4 W. In our measurement, the X-ray tube voltage was kept at 50 kV while the peak X-ray energy was set at 10 keV with an Al/W filter and a 2-mm-diameter brass collimator. The distance between the X-ray source and the X-ray photoconducting device was about 30 cm. The current–voltage measurement of the devices was

performed using a Signotone Micromanipulator S-1160 probe station equipped with a Keithley 4200 Semiconductor Parametric Analyzer. All the experiments were carried out at ambient conditions.

X-ray scintillation detector and imaging. The X-ray scintillation detector was constructed by coating of a PDMS substrate with perovskite QDs (layer thickness of 120 μm), followed by the attachment of a PMT. In a typical procedure, a solution of $CsPbBr_3$ QDs was spin-coated onto the PDMS substrate. The PDMS substrate was then coupled to the PMT for maximized collection of visible photons. For X-ray detection, a range of X-ray dose rates (0.013–278 μGy s^{-1}) was applied by adjusting the current and voltage of the X-ray source. For X-ray imaging, a plastic disk coated with $CsPbBr_3$ nanocrystals was used. A green scarab beetle implanted with a metallic needle was employed as a specimen for X-ray imaging.

In vivo multicolour optical bioimaging. All the animal experiments were performed in compliance with institutional guidelines. Silica-coated perovskite QDs ($CsPbBr_3$, $CsPbBr_{1.5}I_{1.5}$, $CsPbBr_{1.2}I_{1.8}$; 100 μg, 50 μl) dispersed in a phosphate-buffered saline buffer solution were subcutaneously injected into the Balb/c nude mice (age, 4–6 weeks; weight, 18 g). An animal imaging system (Advanced Molecular Imager, Cold Spring Biotech Corp., Shanghai) equipped with an X-ray source was used to carry out in vivo radioluminescence imaging of the mice. The exposure time for in vivo imaging was set at 1 s. For in vivo multicolour optical imaging, optical filters (530 nm, 630 nm and 670 nm) were used to selectively record the X-ray-induced luminescence at different emission wavelengths (Extended Data Fig. 11).

Construction of perovskite-based flat-panel X-ray imaging system. The α-Si photodiode assay backplane was customized for commercial α-Si/CsI:Tl detectors supplied by iRAY Technology Shanghai, Inc. The active area of a photodiode array is 43.0 cm × 43.0 cm, consisting of 3,072 × 3,072 square pixels with a pixel pitch of 139 μm. $CsPbBr_3$ nanocrystals were first dispersed in cyclohexane. We coated the photodiode arrays (8.0 cm × 8.0 cm) with a thin film (thickness, 75 μm) of nanocrystals using a solution-processing method. After evaporation of cyclohexane, an aluminium film (40 μm thick) was added under vacuum, in a packaging process similar to that used in commercial CsI:Tl-based X-ray imaging systems. The aluminium film was used to protect the scintillators from moisture and light soaking. We note that a reflecting layer was coated on the surface of the aluminium film to enhance the light collection into the photodiode elements. The power consumption was 25 W for full-image acquisition and the X-ray source was operated at a voltage of 70 kV. X-ray imaging of electronic circuit boards was acquired with an X-ray exposure of 2.5 mGy s^{-1} for 6 ms, resulting in a dose of 15 μGy. The spatial resolution was determined by measurement of the modulation transfer function.

Radioluminescence analysis using synchrotron radiation. The characterization of the yield of X-ray-induced luminescence near electronic-shell edges was conducted using the synchrotron beamline in the Shanghai Synchrotron Radiation Facility. A thin film of $CsPbBr_3$ nanocrystals was cast onto a PDMS substrate. The X-ray excitation energies were 10–38 keV, and a portable spectrophotometer (Ocean Optics) was used to measure the radioluminescence.

Density functional theory calculation. For the calculation of the projected partial density of states (PDOS), density functional theory (DFT) calculations were carried out. We used the Cambridge Serial Total Energy Package (CASTEP) source code to perform the calculations with the rotation-invariant DFT+U method. In a typical procedure, a simple cubic phase with $Pm\bar{3}m$ symmetrical lattice arrangement was modelled for bulk-phase $CsPbBr_3$. Norm-conserving pseudopotentials of the Cs, Pb and Br atoms were generated by the OPIUM code in the Kleinman–Bylander projector form. A nonlinear partial core correction and a scalar relativistic averaging scheme were used to treat the spin–orbit coupling effect. In particular, we treated the 4*s*, 4*p* and 4*d* states of the Br atoms as valence states, the 5*s*, 5*p* and 5*d* states for Cs atoms, and the 5*d*, 6*s* and 6*p* states for Pb atoms. The Rappe–Rabe–Kaxiras–Joannopoulos method was chosen to optimize the pseudopotentials during electronic minimization, particularly using a blocked-Davidson-scheme matrix diagonalization.

For the calculations of the electronic states in the $CsPbBr_3$ material, we used self-consistent determination for on-site U correction on the localized *p* orbitals of Br sites to correct the on-site Coulomb energy of the electron spurious self-energy. The on-site electronic self-energy and related wavefunction relaxation in the semicore *p*, *d* or *f* orbitals in mixed-valence elements were used to obtain accurate orbital eigenvalues for the electronic structures and transition levels. An ab initio two-way crossover searching calculation was performed by two functionally compiled CASTEP-17 source codes. Using the self-consistent determination, on-site Hubbard U parameters for different orbitals of Br and Pb sites were obtained. Further, a time-dependent DFT calculation was performed with a two-electron-based Tamm–Dancoff approximation imported from the self-consistently corrected ground-state wavefunctions.

Luminescence mechanisms in perovskite nanocrystal scintillator. Two energy-transfer mechanisms for recombination luminescence exist in the perovskite nanocrystal scintillator. One is anisotropic electron and hole transport within the reciprocal Brillouin zone, which leads to a difference between the electron effective mass along different paths and the excitonic binding energy. This difference illustrates the probability that electronic transport within reciprocal band structures is directionally selected for luminescence. Another plausible route is the annihilation of shallow acceptor levels (Pb vacancies), which induces an absence of recombination centres. Such intrinsic lattice defects usually produce low-excitation energy levels compared with the ideal lattice and consequently hinder energy transfer during the light-absorption process.

Anisotropic transport-induced luminescence contrast. Using DFT calculations, the bandgap of a bulk $CsPbBr_3$ crystal was calculated to be about 2.02 eV, whereas in the $CsPbBr_3$ QDs the bandgap increases slightly to 2.22 eV (Fig. 2d). This is because the large surface-to-volume ratio in the $CsPbBr_3$ nanocrystal induces an evident quantum confinement, thus leading to an enlarged vacuum Coulomb barrier for electronic transitions.

We chose reciprocal, highly symmetrical points and lined up two different paths (X→R→M→Γ→R and Γ→R→M) within the Brillouin zone. As shown in the electronic band structure plot in Fig. 2d, the valence band edge and the conduction band edge are located at the same point R(1/2, 1/2, 1/2). Using effective mass theory, the effective mass for electrons and holes was found to be anisotropic along these two directions. In the directions X→R→M→Γ→R and Γ→R→M, the effective masses for electrons were calculated to be $0.03m_0$ and $0.11m_0$ and the effective masses for holes were calculated to be $0.12m_0$ and $0.24m_0$, respectively, where m_0 is the rest mass of the electron. Thus, the Wannier–Mott exciton binding energy and radius are different in these two directions.

By converting the reciprocal Brillouin zone area into a real-space diagram, we found that the Cs site at Γ(0, 0, 0) in the body-centred area is different in these two paths. Point R(1/2, 1/2, 1/2) denotes the position of the Pb site, whereas M(1/2, 1/2, 0) represents the location of the Br site. Owing to the different effective electron masses of the two paths, the path Γ→R→M is energetically favourable to the transport of electrons and holes. By contrast, the X→R→M direction is ruled out because the binding energies are too large to release electrons and holes for recombination. This implies a charge transfer process from the Cs site to the Pb site at the cubic apex point, finally reaching the Br site at the middle point of the cubic edge, namely, through the path Γ→R→M (Extended Data Fig. 7a).

Furthermore, we used an orbital calculation to retrieve the electronic and hole orbitals from the electronic band structure. Our results show that bound electrons stay at the Br sites at a non-bonding state in the p–π orbital level (Extended Data Fig. 7b). Meanwhile, bound holes were found to stay at the Pb site with an *s*-orbital spherical distribution (Extended Data Fig. 7c). The orbital contour plots reveal the localization of electrons and holes at perfect lattices. The stabilized charge state of the body-centred Cs site is Cs^+ because electrons are transferred from the Pb site to the Br site through the ionization of one *s*-state electron.

Intrinsic lattice defects in perovskite nanocrystal scintillators. Intrinsic lattice defects in perovskite nanocrystal scintillators are responsible for both luminescence and the quenching effect. Here we consider the low-energy native defects of a Br vacancy (V_{Br}) and a Pb vacancy (V_{Pb}). For V_{Br}, the absence of one Br atom leaves one electron occupying the empty *p* orbitals of the nearest neighbouring Pb site. Accordingly, localized electronic orbitals were modelled for V_{Br} in the neutral (V^0_{Br}) and singly positive (V^+_{Br}) states (Extended Data Fig. 7d, e). Because the charge bound at the nearby Pb site is positively ionized, the *p* electronic orbitals of the V^+_{Br} site show a transition from the correlated state to a repulsive behaviour between the two neighbouring Pb sites. The PDOS analysis also shows that the electronic level of V_{Br}, which is localized at the bottom of the conduction band edge, serves as a shallow donor. The V_{Pb} lattice defects produce an acceptor trap centre and a spin-polarized state (Extended Data Fig. 7f, g). The singly negative state of a Pb site (V^-_{Pb}) with one electron already captured could partially passivate the acceptor trap site with weakened charge localization. The process of local geometrical relaxation on different charge states of V_{Pb} indicates that the Cs^+ sites centrosymmetrically move towards the V_{Pb} centre (Extended Data Fig. 7h). Upon the occurrence of a strong clustering effect to form V^{2-}_{Pb} near the V_{Pb} site, the electronically active acceptor trap centre can be completely terminated. In this case, formation of local Cs–Br motifs is possible. We also considered V_{Cs} and V^-_{Cs} sites in the lattice and found no effects on the electronic properties of the host lattice. Accordingly, V_{Br} and V_{Pb} produce the electronic and hole levels for luminescence recombination in the form of photon emission (Extended Data Fig. 7i).

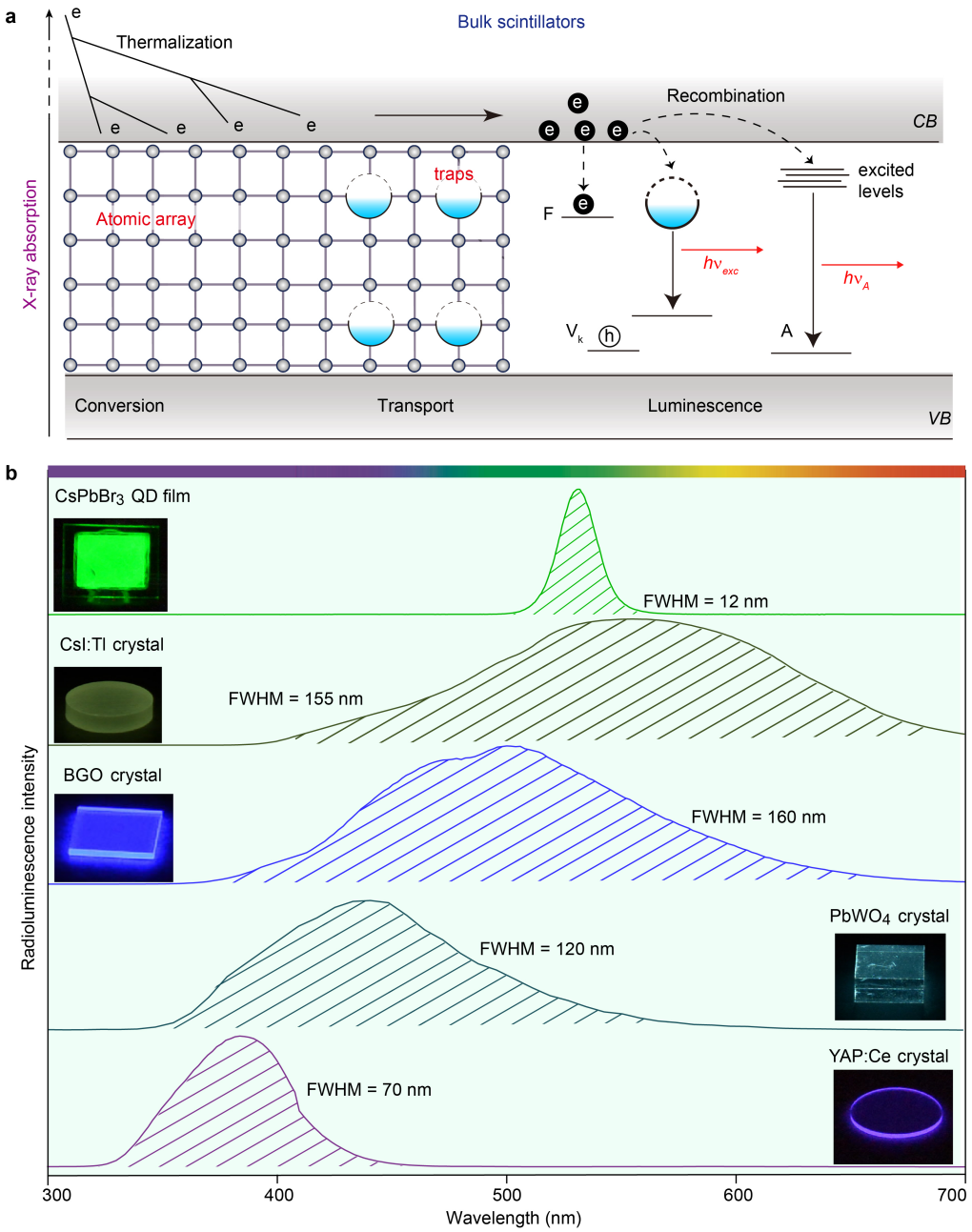
Further, we calculated the excited energy and thermodynamic transition levels in the $CsPbBr_3$ QDs. From the internal bulk lattice to the surface region, the dimensions of the material decrease but its surface-to-volume radio is increased. The electronic donor trap levels experience a transition from a localized state below the conduction band edge to a delocalized state in the conduction band (Extended Data Fig. 7j). During the process of radiation ionization and release, the number of bound electrons is increased accordingly. By contrast, with a decreased dimension of the host lattice, the trapping ability of the acceptor is decreased where the hole level shifts from a delocalized state in the valence band to a localized state above

the valence band. Therefore, the quenching effect in the bulk $CsPbBr_3$ materials for luminescence recombination is caused by annihilation of a hole level that is deeply buried in the valence band. The structural transition from the QD to the bulk form occurs from the surface to the bulk, thus the hole level is annihilated.

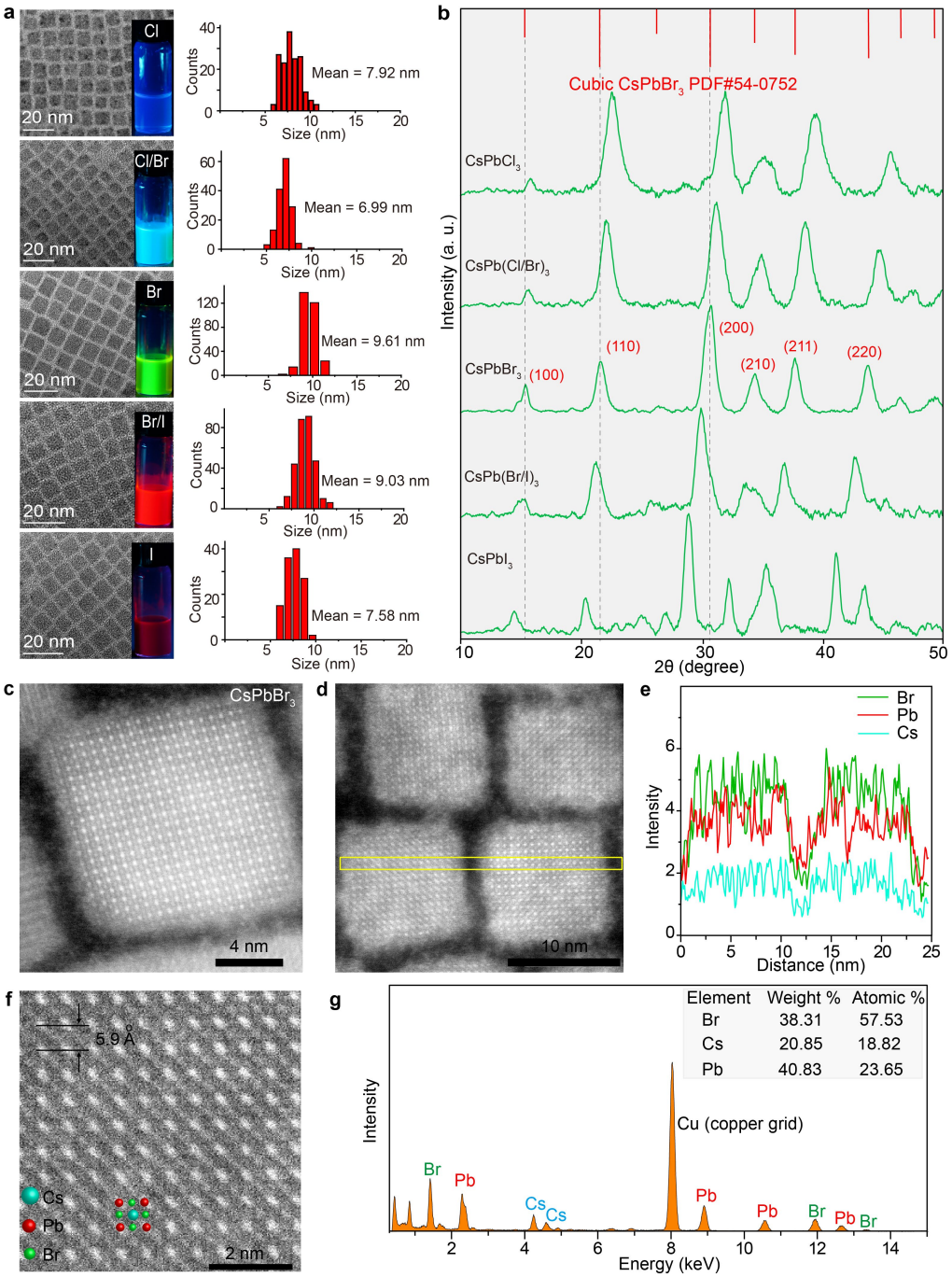
Intrinsic quantum confinement in $CsPbBr_3$ nanocrystals. The intrinsic effect of quantum confinement in $CsPbBr_3$ nanocrystals was examined by additional theoretical study of their surface electronic properties. In a typical procedure, we first built a simplified model of the $CsPbBr_3$ structure composed of 293 atoms with a particle size of 12.06 Å, namely, a lattice group ($6 \times 6 \times 6$) truncated from bulk $CsPbBr_3$ crystal, using the radial coordinated structural formation program (RCSFP) (Extended Data Fig. 8a). DFT calculations yield that the orbital contour plots of the $CsPbBr_3$ QD show the highest occupied molecular orbital (HOMO), the lowest unoccupied molecular orbital (LUMO) and surface-vacancy-induced Coulomb trapping (SVIC) states (Extended Data Fig. 8b). The electronic structures show that the SVIC state is formed owing to unsaturated *p* orbitals of surface Br sites and is electronically localized at the apex corner regions of the QD. Additionally, the PDOS of the $CsPbBr_3$ QD indicates that such SVIC sites are mainly distributed near the Fermi level, beyond the valence band maximum, thus exhibiting a hole-like feature and being strongly confined by the LUMO orbitals (Extended Data Fig. 8c). This leads to the suppression of the long-distance transport of electron–hole pairs across the particle surface or between particles. Furthermore, a model was built to perform a simulation of the energetic evolution on the surface of the QDs as a function of particle distance (Extended Data Fig. 8d). To investigate surface confinement, we further calculated the relative energy level of the SVIC state as a function of particle distance (Extended Data Fig. 8d). It is obvious that at a distance of 12.06 Å from the particle surface, the SVIC energy implies a strong hole-like confinement, merely 0.056 eV above the valence band maximum. Our results suggest that the mean confinement path of electronic transport within the lattice is approximately 10.32 Å (Fig. 2f). Indeed, the intrinsic energetics on the surface of the QDs is reasonable for energy confinement of the thermalized low-energy excitons inside the nanocrystal, resulting in a high yield of X-ray scintillation light.

Data availability. The data that support the findings of this study are available from the corresponding authors upon reasonable request.

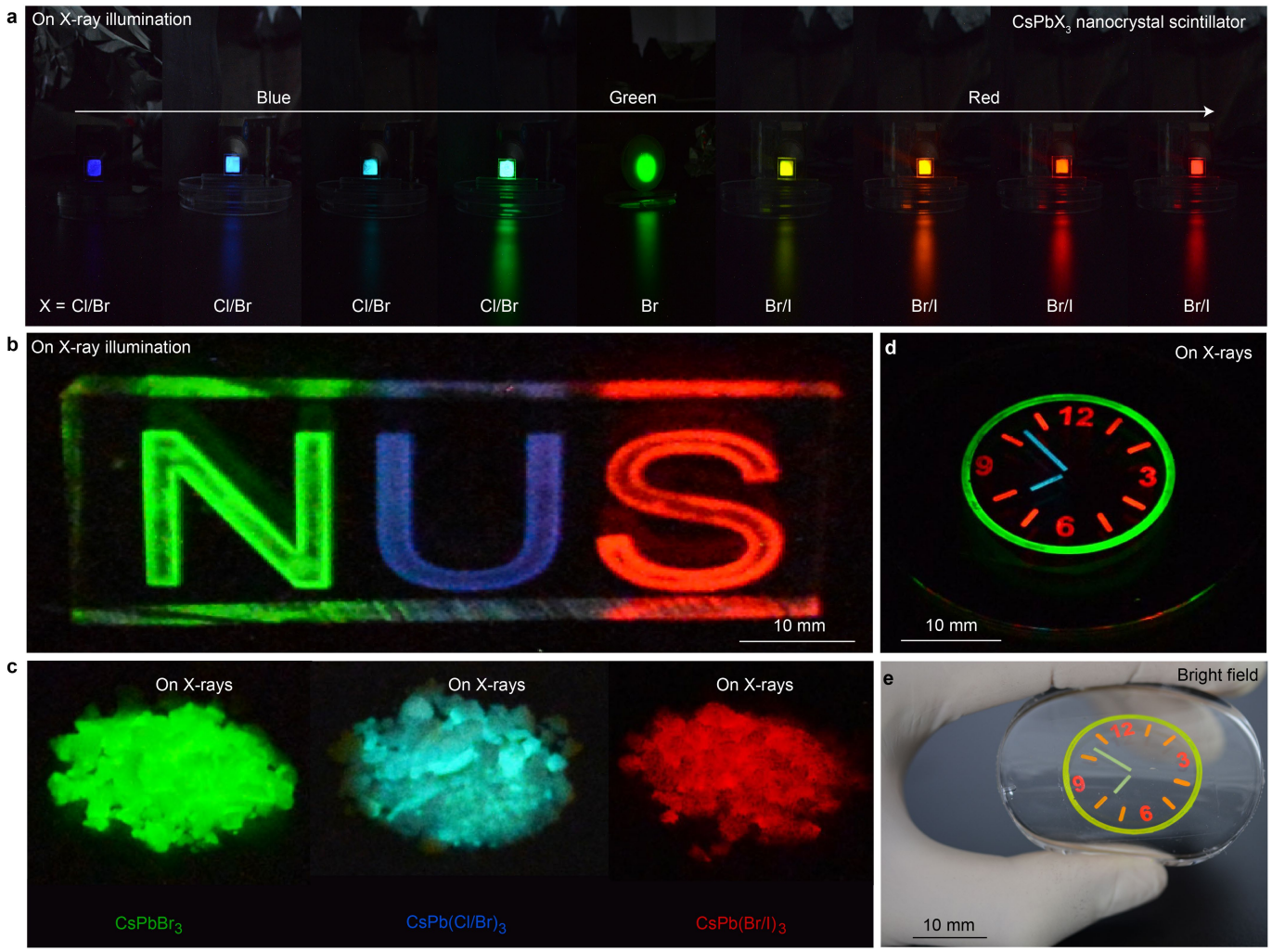
34. Saidaminov, M. I. et al. Inorganic lead halide perovskite single crystals: phase-selective low-temperature growth, carrier transport properties, and self-powered photodetection. *Adv. Opt. Mater.* **5**, 1600704 (2017).
35. Huang, S. et al. Enhancing the stability of $CH_3NH_3PbBr_3$ quantum dots by embedding in silica spheres derived from tetramethyl orthosilicate in "waterless" toluene. *J. Am. Chem. Soc.* **138**, 5749–5752 (2016).
36. Nikl, M. Scintillation detectors for X-rays. *Meas. Sci. Technol.* **17**, 37–54 (2006).
37. Miyata, A. et al. Direct measurement of the exciton binding energy and effective masses for charge carriers in organic–inorganic tri-halide perovskites. *Nat. Photon.* **11**, 582–587 (2015).



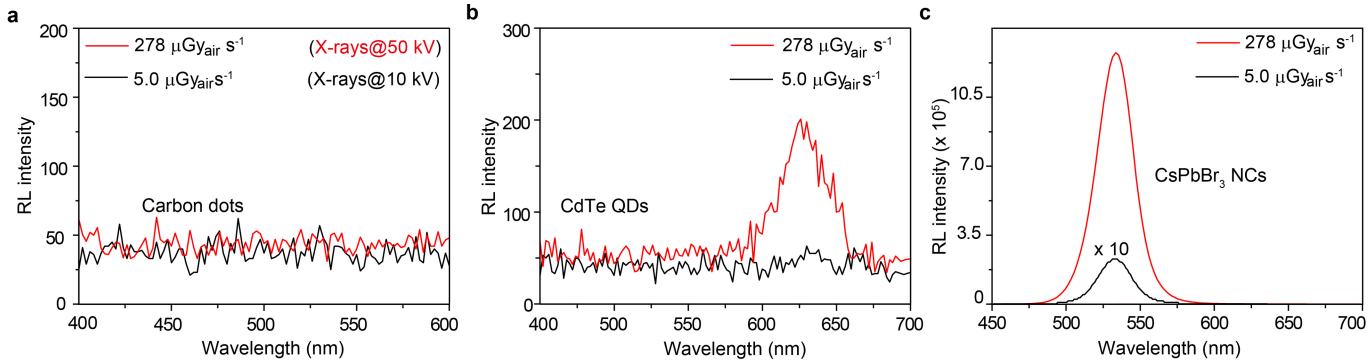
Extended Data Fig. 1 | Schematic representation of scintillation mechanism and X-ray-induced luminescence in bulk inorganic scintillators. a, In bulk inorganic materials, X-ray photons absorbed by the lattice atoms can generate hot charge carriers, followed by exciton thermalization and subsequent transport to defect sites or activator centres ('traps'), where radiative emission occurs. VB, valence band; CB, conduction band; V_k, self-trapped hole; F, Farbe centre; ν_{exc}, photon frequency during excitonic luminescence; ν_A, photon frequency during activator (A) luminescence. **b**, Scintillation properties of $CsPbBr_3$ QDs and commercial bulk inorganic materials under X-ray excitation. The full-width at half-maximum (FWHM) represents the spectral width of the scintillation spectra of the crystal. The insets show photographs of the corresponding samples acquired under X-ray excitation. BGO, $Bi_4Ge_3O_{12}$.

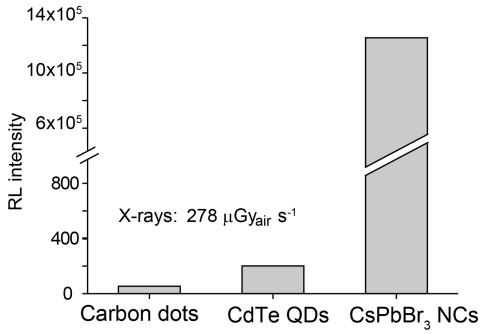


Extended Data Fig. 2 | Physical characterization of as-synthesized perovskite QDs. a, TEM images of the as-prepared cubic-phased nanocrystals (left) and the corresponding size distribution of the nanocrystals (right). The samples are $CsPbCl_3$, $CsPb(Cl/Br)_3$, $CsPbBr_3$, $CsPb(Br/I)_3$ and $CsPbI_3$ nanocrystals (from top to bottom). Insets are images of perovskite nanocrystals dispersed in cyclohexane, recorded under 365-nm ultraviolet light excitation. **b**, Powder X-ray diffraction patterns for typical ternary and mixed-halide $CsPbX_3$ (X = Cl, Br or I) nanocrystals. All peaks are indexed in accordance with the cubic-phased $CsPbBr_3$ structure (Joint Committee on Powder Diffraction Standards file (PDF) number 54-0752). **c**, **d**, Dark-field scanning transmission electron micrographs (STEM; JEM-F200HR) of $CsPbBr_3$ nanocrystals. **e**, Elemental mapping of Br, Pb and Cs for the nanocrystals, obtained from the area marked by the rectangular box in **d** using energy-dispersive X-ray spectroscopy. **f**, Atomically resolved dark-field STEM image of a single $CsPbBr_3$ nanocrystal, showing Cs and Pb lattice atoms. **g**, Energy-dispersive X-ray spectrum of the as-prepared $CsPbBr_3$ perovskite nanocrystals, confirming the stoichiometric composition of the $CsPbBr_3$ nanocrystals. We note that strong Cu signals come from the TEM copper grid.



Extended Data Fig. 3 | Multicolour-emitting perovskite QD scintillators upon X-ray irradiation. a, Demonstration of X-ray-induced luminescence modulation using $CsPbX_3$ QDs of different compositions (X = Cl, Br or I). **b**, Multicolour X-ray scintillation from $CsPb(Cl/Br)_3$, $CsPbBr_3$ and $CsPb(Br/I)_3$ nanocrystals cast on a PDMS substrate. The X-ray dose rate is 278 μGy s^{-1}. **c**, Typical photographs of radioluminescence from $CsPb(Cl/Br)_3$, $CsPbBr_3$ and $CsPb(Br/I)_3$ QDs under X-ray excitation. **d**, **e**, Multicolour visualization of as-developed perovskite QD scintillators using X-rays (**d**) and the corresponding bright-field image (**e**).

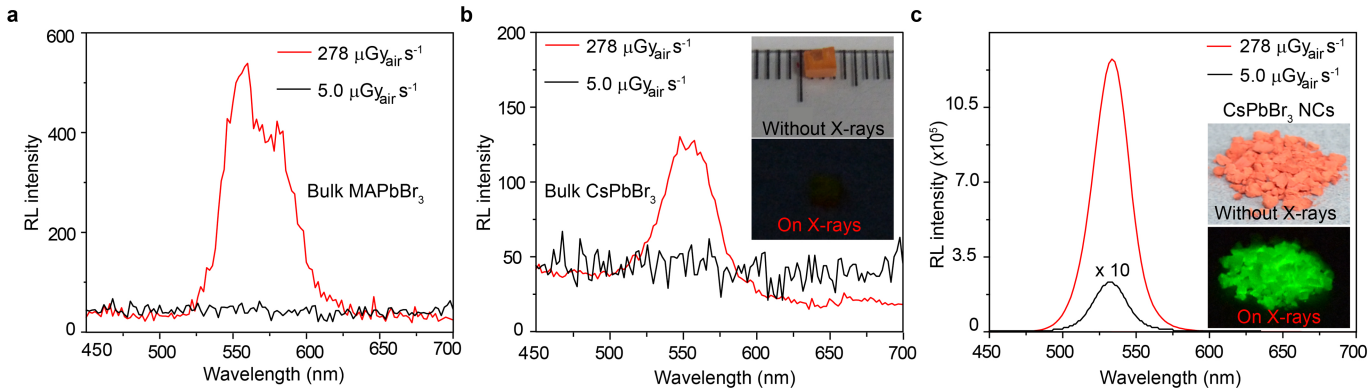


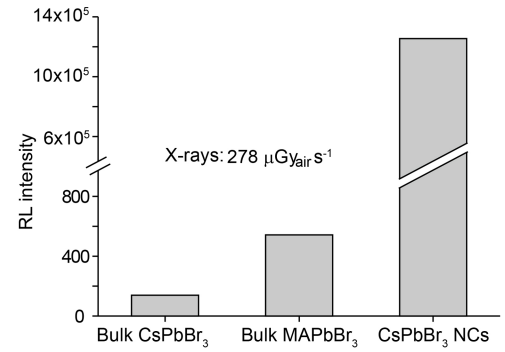


e

Chem. Compos.	Emission Max. [nm]	X-ray-excited Luminescence	K-edge Max. [keV]
$CsPbBr_3$ NCs	532	Strong	88.0
CdTe QDs	626	Very weak	31.8
Carbon dots	526	-	0.285

Extended Data Fig. 4 | Comparison of X-ray-induced luminescence intensity for QDs. a–c, Radioluminescence spectra of carbon dots (**a**), CdTe QDs (**b**) and $CsPbBr_3$ nanocrystals (NCs; **c**) under X-ray excitation at 5.0 μGy s^{-1} and 278 μGy s^{-1}. **d**, Comparison of radioluminescence intensity for the as-prepared carbon dots, CdTe QDs and $CsPbBr_3$ nanocrystals under 278 μGy s^{-1} X-ray excitation. **e**, Scintillation performance and maximal K-edge energy of lead halide perovskite nanocrystals, carbon dots and CdTe QDs.

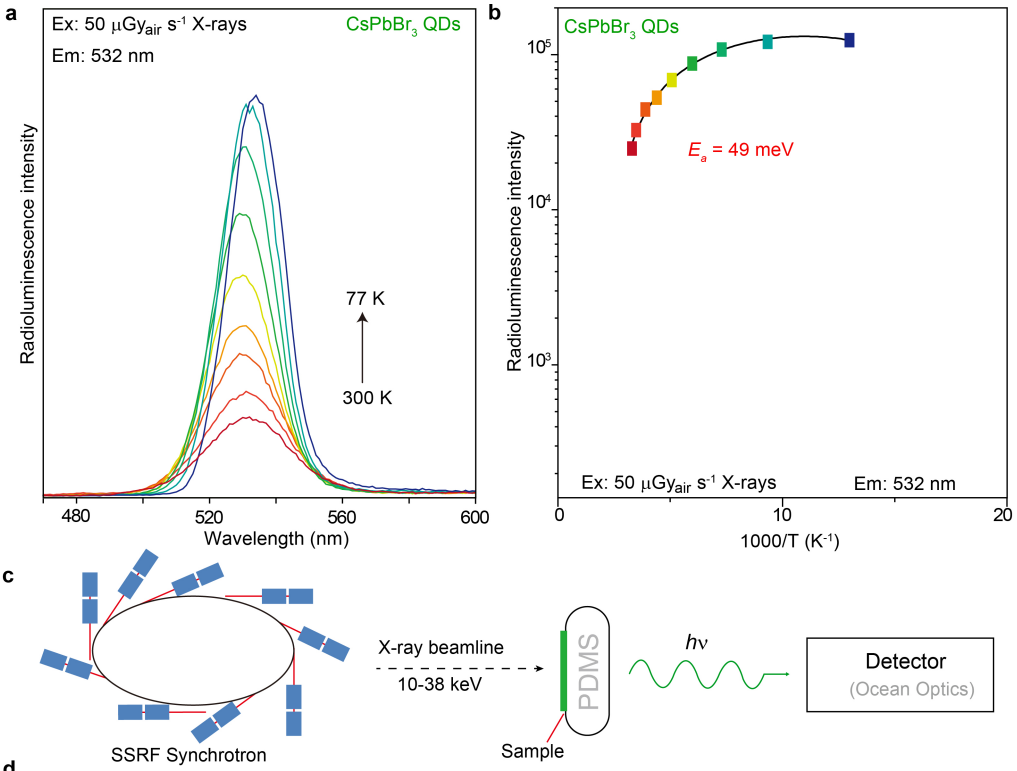




e

Chem. Compos.	PLQY (%)	Binding Energy (meV)	X-ray-excited Luminescence
$CsPbBr_3$ nanocrystals	58.6	45-120	Strong
Bulk $CsPbBr_3$	< 0.1	35-40	Very weak
Bulk $MAPbBr_3$	< 0.1	< 15	Very weak

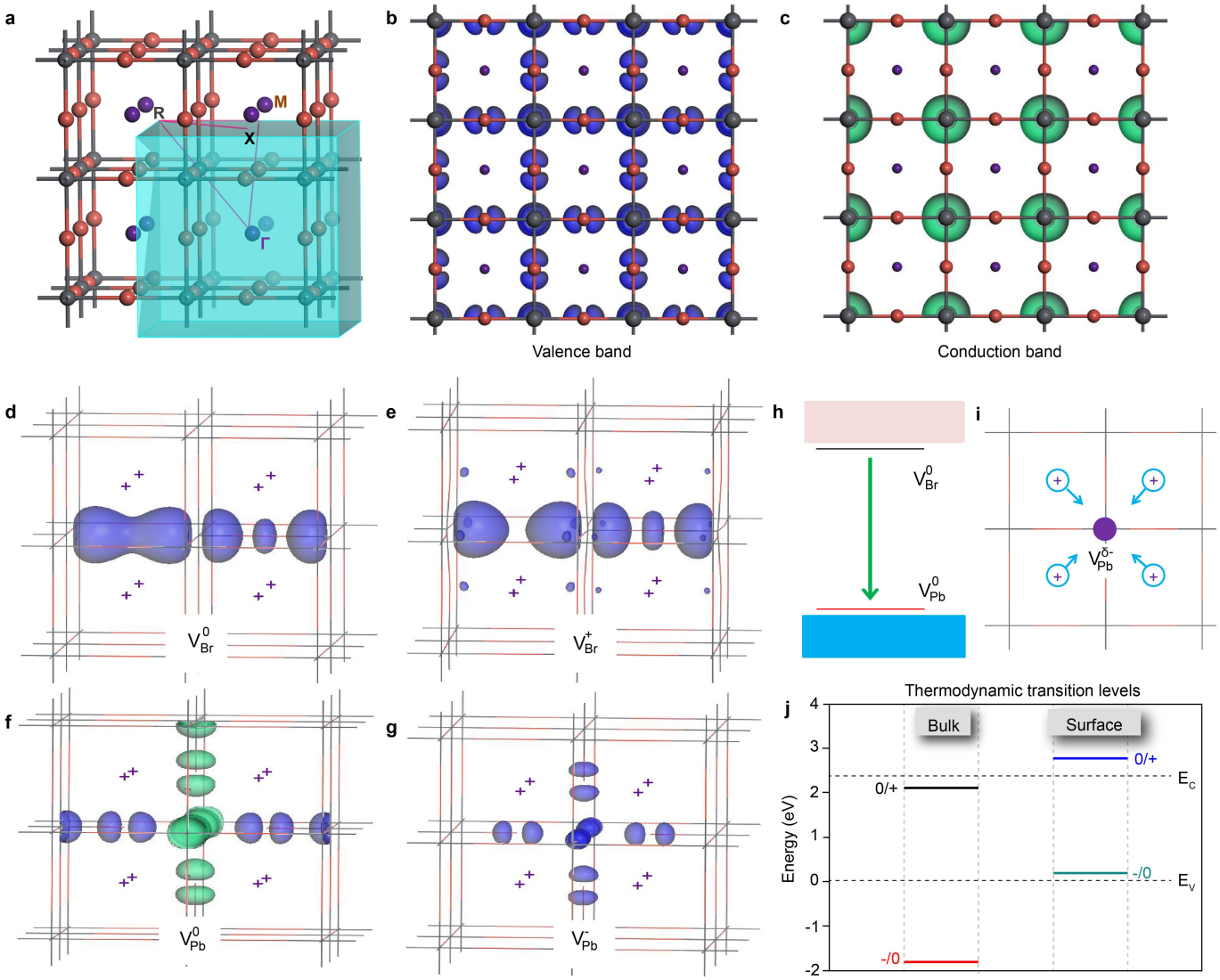
Extended Data Fig. 5 | Comparison of X-ray-induced luminescence for lead halide perovskite materials. a–**c**, Radioluminescence spectra of bulk single-crystal $CH_3NH_3PbBr_3$ (**a**) and $CsPbBr_3$ (**b**) and of $CsPbBr_3$ nanocrystals (**c**) under X-ray excitation at 5.0 μGy s^{-1} and 278 μGy s^{-1}. The insets are photographs of $CsPbBr_3$ bulk single crystal (**b**) and nanocrystal powders (**c**) taken under ambient light (top) and X-ray illumination (bottom). The X-ray dose used for the experiments was 278 μGy s^{-1}. **d**, Comparison of radioluminescence intensity for three types of perovskite material under 278 μGy s^{-1} X-ray excitation. **e**, PLQY and exciton binding energy of perovskite materials at 300 K[31,37]. We note that the thermal energy at 300 K is $k_BT \approx 25$ meV. Nanocrystalline perovskites are highly luminescent materials, whereas bulk perovskite crystals are most suitable for the generation of free charge carriers owing to their low exciton binding energy.



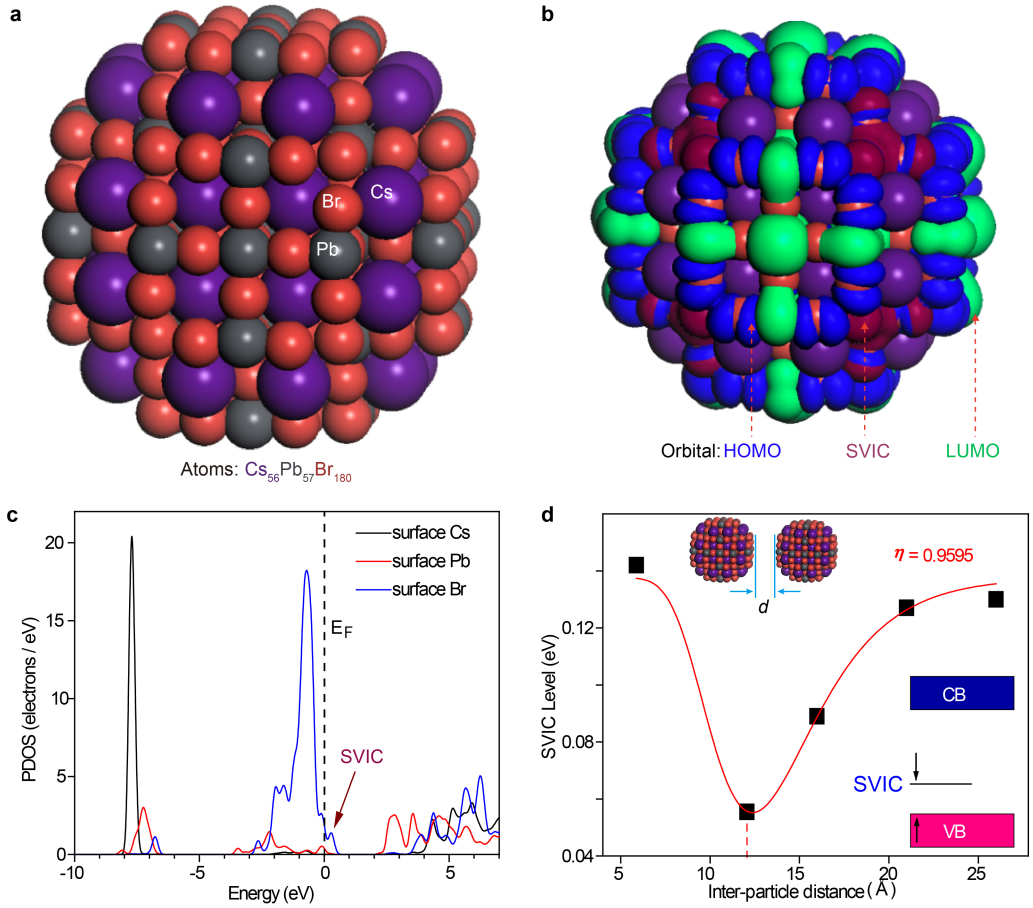
d

Edge	Photon Energy (keV)		
Pb-L_3	13.04	13.25	13.47
Br-K	13.47	15.00	15.20
Pb-L_2	15.20	15.53	15.86
Pb-L_1	20.00	30.00	35.98
Cs-K	35.98	40.00	
Pb-K	88.00	100.0	

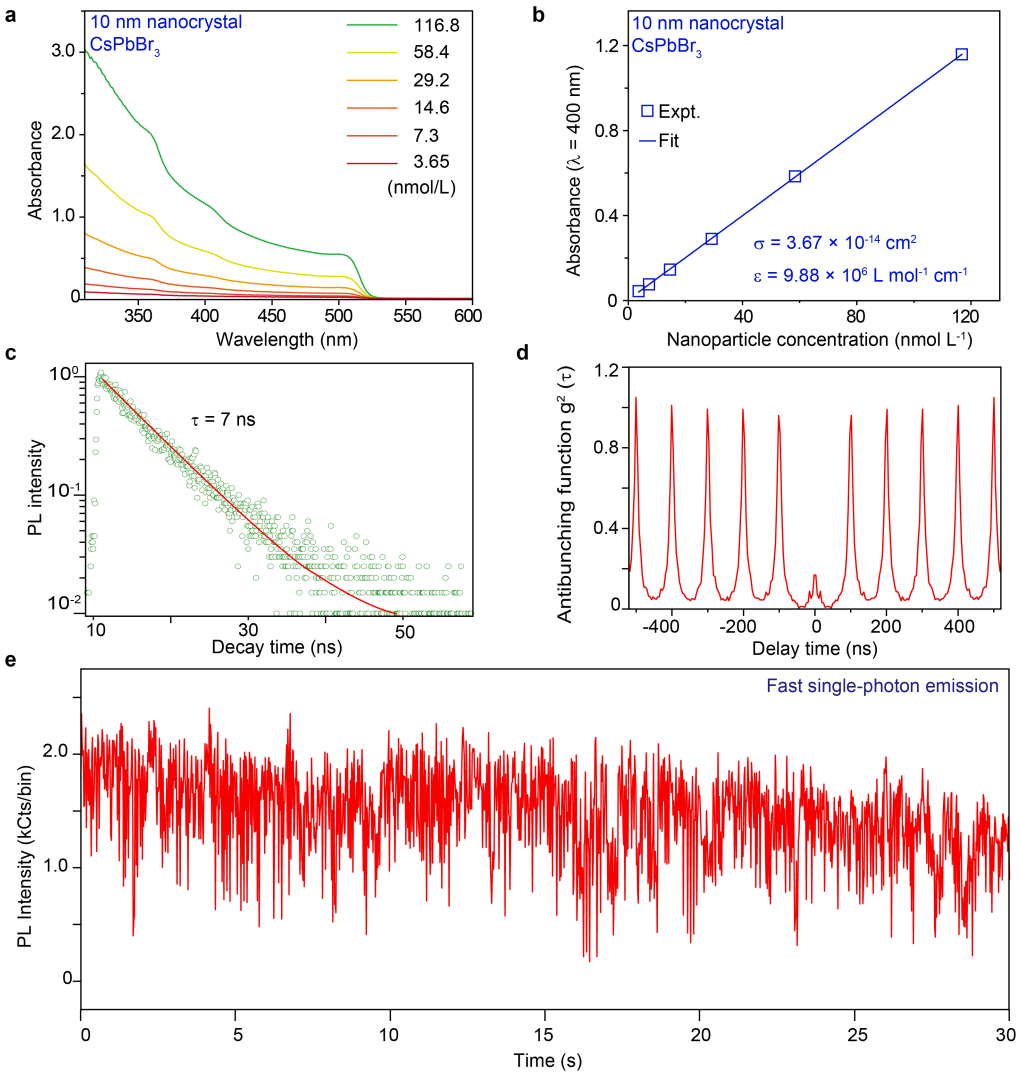
Extended Data Fig. 6 | Measurement of exciton binding energy and synchrotron-radiation-induced radioluminescence of $CsPbBr_3$ nanocrystals. a, Temperature-dependent scintillation spectra of the $CsPbBr_3$ nanocrystals at 77–300 K under X-ray illumination at 50 μGy s^{-1}. **b**, Arrhenius plot of the X-ray-induced luminescence intensities of the $CsPbBr_3$ nanocrystals at 532 nm. **c**, Experimental setup for the synchrotron-radiation-induced radioluminescence measurements at the X-ray beamline of the Shanghai Synchrotron Radiation Facility (SSRF). The X-ray energy is 10–38 keV. **d**, The electronic edge energies for Pb L, Cs K and Br K.



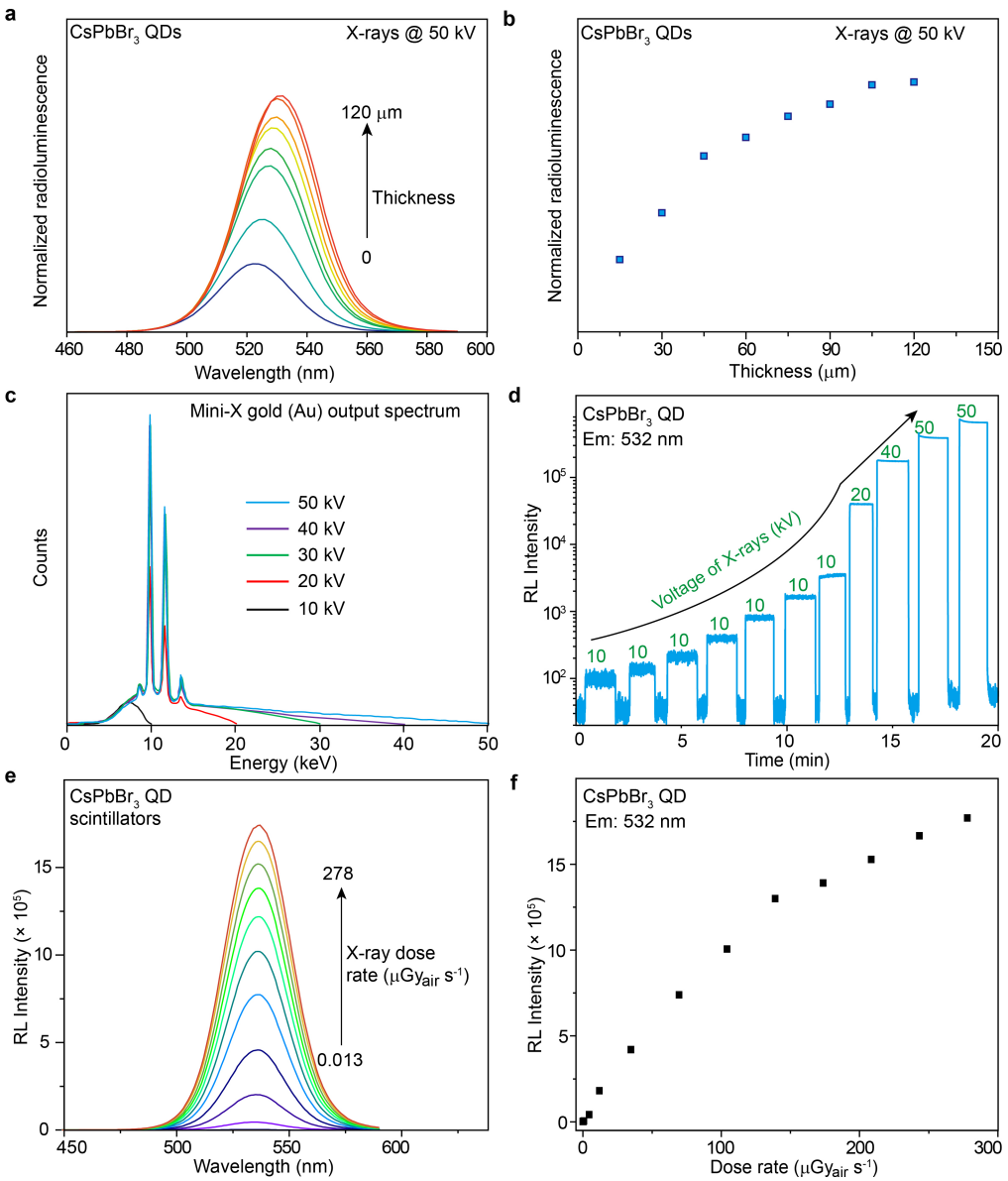
Extended Data Fig. 7 | Electronic structure and scintillation mechanism of the $CsPbBr_3$ nanocrystal along selected reciprocal high-symmetry points within the Brillouin zone. a, The Brillouin zone of the cubic-phased crystal lattice in $CsPbBr_3$, calculated using relativistic corrections. Γ, M, R and X denote high-symmetry points within the reciprocal space (blue). **b**, **c**, Calculated electron density associated with the valence band (**b**) and the conduction band (**c**) of the cubic $CsPbBr_3$. We note that the halide ions contribute to the changes of the bandgap in the perovskite nanocrystal through the influence of the valence orbitals. Cs, purple; Pb, grey; Br, orange; electron orbital, blue; hole orbital, green. **d**–**g**, Localized electronic and hole levels for V_{Br} and V_{Pb} at different charge states. **h**, **i**, Schematic diagram of energy transfer for radioluminescence induced by intrinsic lattice defects and the quenching effect caused by evident ion movement (Cs^+). $V_{Pb}^{\delta-}$ denotes a Pb vacancy with small ($|\delta| < 1$) charge transfer. **j**, Thermodynamic transition levels of the perovskite nanocrystal. E_c, maximum valence band energy; E_v, minimum conduction band energy.



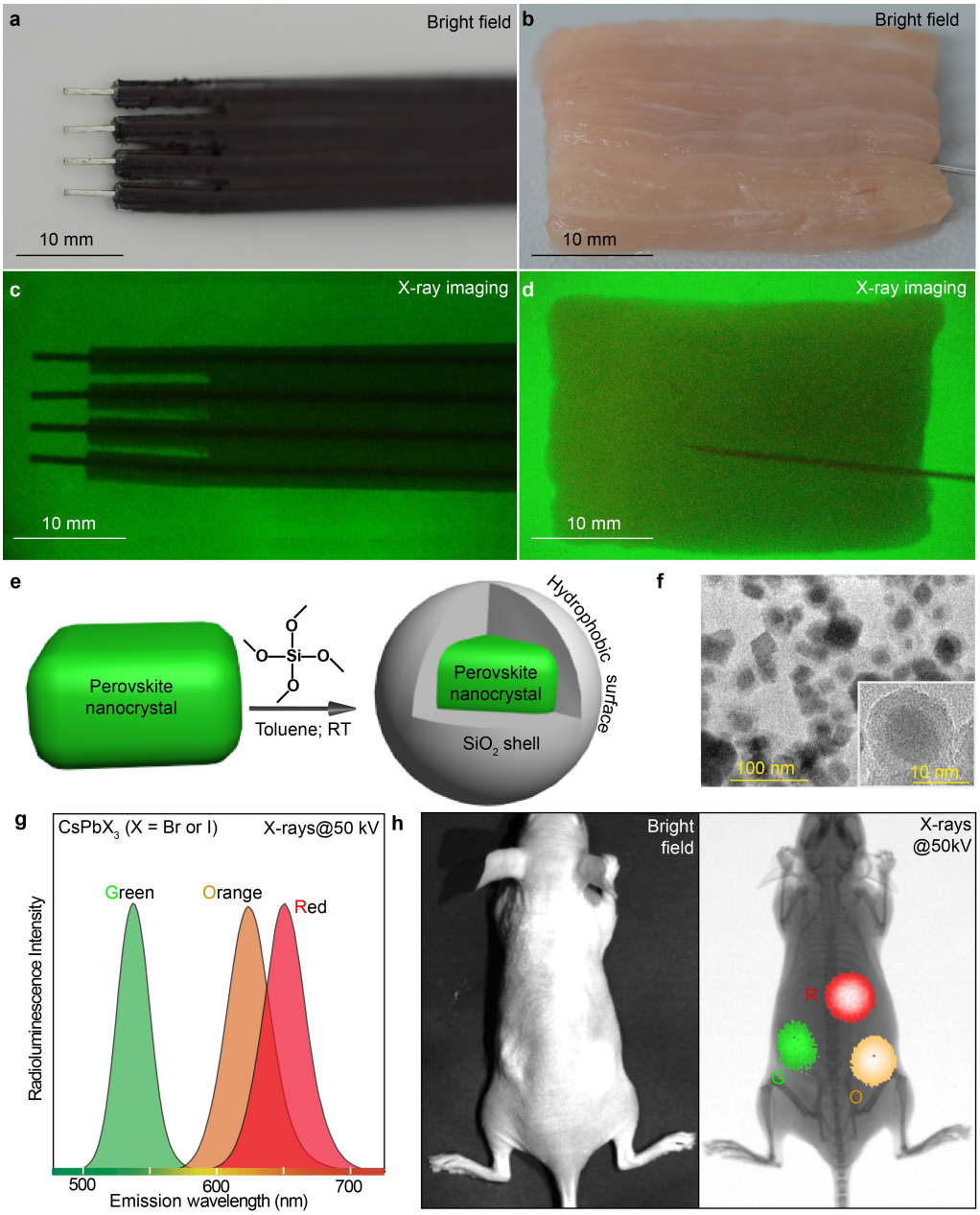
Extended Data Fig. 8 | Theoretical studies of the surface electronic properties of $CsPbBr_3$ nanocrystals. a, Simplified model of $CsPbBr_3$ structure composed of 293 atoms with a particle size of 12.06 Å. **b**, Calculated orbital contour plots of a $CsPbBr_3$ nanocrystal, showing the HOMO (blue), LUMO (green) and SVIC trapping (red) states. The SVIC states are formed owing to unsaturated *p* orbitals of surface Br sites. **c**, PDOS of the $CsPbBr_3$ QD. The SVIC state is located near the Fermi level (E_F). **d**, The relative energy of the SVIC state to the valence band (VB) maximum as a function of inter-particle distance, *d*. The red line is calculated by fitting with the Gaussian distribution function. The top inset shows the simulation model used to calculate the energy evolution of the $CsPbBr_3$ nanocrystal as a function of particle distance. CB, conduction band.



Extended Data Fig. 9 | Characterization of absorption cross-section and transient luminescence spectra of the $CsPbBr_3$ nanocrystals. **a**, Absorption spectra of $CsPbBr_3$ nanocrystals dispersed in cyclohexane at different concentrations. **b**, Absorption as a function of concentration for $CsPbBr_3$ nanocrystals. The molar extinction coefficient, ε, was determined by applying the Beer–Lambert law $A = \varepsilon cL$, where A is the absorbance, c is the molar concentration (mol l^{-1}) and L is the optical path length (1 cm) through the sample. The absorption cross-section, σ, was determined by $\varepsilon = N_A\sigma/(1{,}000 \times \ln 10)$, where N_A is Avogadro's number. **c**, Photoluminescence (PL) lifetime of a single $CsPbBr_3$ perovskite nanocrystal. **d**, Second-order correlation function, $g^2(\tau)$, of the nanocrystal. The value $g^2(0) = 0.16$ confirms the single-quantum-emitter nature of the photon emission. **e**, Fluorescence intermittency trace recorded for a single $CsPbBr_3$ perovskite nanocrystal. The recorded photoluminescence intensity reaches more than 2,000 counts per 20-ms bin.



Extended Data Fig. 10 | Performance of X-ray detection system using $CsPbBr_3$ nanocrystal scintillators. a, Normalized radioluminescence intensity of $CsPbBr_3$ nanocrystals of different thickness under X-ray excitation at a voltage of 50 kV. **b**, Normalized radioluminescence intensity as a function of perovskite nanocrystal film thickness. **c**, X-ray output spectra recorded at 10, 20, 30, 40 and 50 kV. **d**, Kinetic measurement of radioluminescence intensity at 532 nm in response to an X-ray dose rate of 0.013–278 μGy s^{-1}. **e**, Emission spectra of $CsPbBr_3$ nanocrystal scintillator in response to an X-ray dose rate of 0.013–278 μGy s^{-1}. **f**, Radioluminescence (RL) intensity as a function of the X-ray dose rate shown in **e**.



Extended Data Fig. 11 | Direct X-ray imaging and multiplexed labelling for in vivo optical imaging using perovskite nanocrystal scintillators. a–d, A flexible flat cable (**a**) and needle-implanted pork tissue (**b**) were imaged with bright-field and X-ray imaging (**c**, **d**). We note that the $CsPbBr_3$ nanocrystal scintillator platform shown in Fig. 3d was used for the X-ray phase contrast imaging. In both cases, the X-ray images clearly reveal the presence of metallic wires embedded in the cable and pork tissue. **e**, Synthesis of $CsPbBr_3/SiO_2$ core-shell nanoparticles with a hydrophobic surface for protection against moisture. RT, room temperature. **f**, TEM image of the as-prepared $CsPbBr_3/SiO_2$ nanoparticles. **g**, Multicolour luminescence spectra of the perovskite nanocrystals under X-ray irradiation at a voltage of 50 kV. The materials' compositions are $CsPbBr_3$, $CsPbBr_{1.5}I_{1.5}$ and $CsPbBr_{1.2}I_{1.8}$ for green (G), orange (O) and red (R) emissions, respectively. We note that the risk of lead toxicity must be considered during experimentation. **h**, Bright-field and multicolour luminescent in vivo imaging in mice under X-ray excitation at a voltage of 50 kV. The X-ray-induced luminescence was recorded by a charge-coupled-device camera equipped with three optical filters at 530 nm, 630 nm and 670 nm.

Extended Data Table 1 | Scintillation characteristics for different materials

Chemical Comp.	Emission Max. (nm)	FWHM (nm)	Thickness (mm)	Decay Time (ns)	Synthetic Temp. (°C)
$CsPbX_3$ QDs	410-700	12-40	0.1	44.6	~160
CsI:Tl	565	155	5.0	1000	~1700
$Bi_4Ge_3O_{12}$	490	160	3.0	300	~1700
$YAlO_3$:Ce	380	55	0.6	10^2-10^4	~1700
$PbWO_4$	440	120	5.0	10^2-10^4	~1700

Extended Data Table 2 | Properties of perovskite nanocrystals and bulk crystals used for X-ray detection

Chemical component	Material type	Electric response	Lumines. response	Multicolor scintillation	Detection limit ($\mu Gy_{air}\ s^{-1}$)	Imaging method	Remarks
$CsPbX_3$ nanocrystals (X=Cl, Br, I)	Scintillator	Low	Highly sensitive	Yes	0.013	Photon-to-photon conversion	*This work*[a]
$MAPbI_3$ thin film	Semiconductor	Good	Not reported	No	~1000	Photon-to-current conversion	*Nature Photon.* **2015**, *9*, 444.
$MAPbBr_3$ single crystal	Semiconductor	Excellent	Not reported	No	0.5	-	*Nat. Photon.* **2016**, *10*, 333.
$MAPbBr_3$ single crystal/Si substrate	Semiconductor	Excellent	Not reported	No	0.036	Photon-to-current conversion	*Nature Photon.* **2017**, *11*, 315.
$MAPbI_3$ microcrystal	Semiconductor	Excellent	Not reported	No	48	-	*Nature Photon.* **2017**, *11*, 436.
$Cs_2AgBiBr_6$ single crystal	Semiconductor	Excellent	Not reported	No	0.0597	-	*Nature Photon.* **2017**, *11*, 726.
$MAPbI_3$ polycrystalline	Semiconductor	Excellent	Not reported	No	-	Photon-to-current conversion	*Nature* **2017**, *550*, 87.
$MAPbBr_3$ nanocrystals	Scintillator	-	Yes	-	-	Photon-to-photon conversion	*This work*[b]

Data are from this work and from refs [12–16,18]. MA, CH_3NH_3.

[a]We note that scintillators are a special class of luminescence materials that have been most widely used for radiation detection by converting high-energy X-ray photons into visible light—a sensing process that is different in semiconductor materials, where the dominant process is photon-to-current conversion. The as-developed perovskite nanocrystal scintillators are solution-processable and can respond to X-rays with multicolour output, which can be readily recorded by a common digital camera.

[b]We note that the organic–inorganic hybrid QDs are less photostable than their purely inorganic counterparts and their synthesis is relatively difficult to scale up owing to the need for stringent control over the reaction conditions.

LETTER

https://doi.org/10.1038/s41586-018-0464-9

A Brownian quasi-crystal of pre-assembled colloidal Penrose tiles

Po-Yuan Wang[1,2] & Thomas G. Mason[1,3]*

Penrose's pentagonal P2 quasi-crystal[1–4] is a beautiful, hierarchically organized multiscale structure in which kite- and dart-shaped tiles are arranged into local motifs, such as pentagonal stars, which are in turn arranged into various close-packed superstructural patterns that become increasingly complex at larger length scales. Although certain types of quasi-periodic structure have been observed in hard and soft matter, such structures are difficult to engineer, especially over large areas, because generating the necessary, highly specific interactions between constituent building blocks is challenging. Previously reported soft-matter quasi-crystals of dendrimers[5], triblock copolymers[6], nanoparticles[7] and polymeric micelles[8] have been limited to 12- or 18-fold symmetries. Because routes for self-assembling complex colloidal building blocks[9–11] into low-defect dynamic superstructures remain limited[12], alternative methods, such as using optical and directed assembly, are being explored[13,14]. Holographic laser tweezers[15] and optical standing waves[16] have been used to hold microspheres in local quasi-crystalline arrangements, and magnetic microspheres of two different sizes have been assembled into local five-fold-symmetric quasi-crystalline arrangements in two dimensions[17]. But a Penrose quasi-crystal of mobile colloidal tiles has hitherto not been fabricated over large areas. Here we report such a quasi-crystal in two dimensions, created using a highly parallelizable method of lithographic printing and subsequent release of pre-assembled kite- and dart-shaped tiles into a solution–dispersion containing a depletion agent. After release, the positions and orientations of the tiles within the quasi-crystal can fluctuate, and these tiles undergo random, Brownian motion in the monolayer owing to frequent collisions between neighbouring tiles, even after the system reaches equilibrium. Using optical microscopy, we study both the equilibrium fluctuations of the system at high tile densities and also the 'melting' of the pattern as the tile density is lowered. At high tile densities we find signatures of a five-fold pentatic liquid quasi-crystalline phase, analogous to a six-fold hexatic liquid crystal. Our fabrication approach is applicable to tiles of different sizes and shapes, and with different initial positions and orientations, enabling the creation of two-dimensional quasi-crystalline systems (and other systems that possess multiscale complexity at high tile densities) beyond those of current self- or directed-assembly methods[18–20]. We anticipate that our approach for generating lithographically pre-assembled monolayers could be extended to create three-dimensional Brownian systems of fluctuating particles with custom-designed shapes through holographic lithography[21,22] or stereolithography[23].

To make lithographically pre-assembled monolayers (litho-PAMs; see Methods), we combine computer-aided design software and lithography to fabricate, position and orient many shape-designed colloidal particles in a desired complex initial configuration. After lithographic printing using an optical stepper and development, we obtain a pre-assembled static set of discrete prismatic polymeric particles, each approximately 2 μm thick and composed of cross-linked epoxy SU-8 photoresist, attached to a thin (approximately 10 nm) layer of a release material (Omnicoat) on a smooth glass wafer (Fig. 1a, Extended Data Fig. 1). The particles (or 'tiles') are shaped like kites (convex quadrilateral prisms) or darts (non-convex quadrilateral prisms) and are arranged in a Penrose P2 quasi-crystalline pattern, as confirmed by optical and scanning electron micrographs (Fig. 1b), at a tile area fraction (ratio of tile area to total monolayer area) of $\phi_A \approx 0.78$. We enclose the printed region in polydimethylsiloxane (PDMS) elastomeric walls and add an aqueous release solution–dispersion (RSD) that we custom-formulated to maintain an intact monolayer of fluctuating tiles that have nearly hard in-plane interactions (that is, the tiles in the monolayer are non-interacting except upon collisional contact, when they repel impulsively, ensuring that tile overlap is forbidden). The RSD contains a base (tetramethylammonium hydroxide) to dissolve the Omnicoat, sodium dodecyl sulfate surfactant to prevent the SU-8 tiles that are released from aggregating or sticking to the glass substrate, and a depletion agent (such as anionic polystyrene nanospheres) to create anisotropic roughness-controlled depletion attractions[24,25] that maintain the monolayer (see Methods and Supplementary Methods for details).

After adding the RSD to the PDMS well (Fig. 1c, Extended Data Fig. 1), we image the tiles as they release using an inverted optical microscope. The previously static tiles begin to undergo Brownian motion, colliding frequently with neighbouring tiles as the system equilibrates at fixed ϕ_A (Fig. 1d). Time-lapse digital microphotography yields high-resolution videos of the release kinetics (Supplementary Video 1). By digitally subtracting the initial image before release from images at later times t after adding the RSD, we determine the fraction of tiles that are released, $P_r(t)$ (Fig. 1e). The observed release profile is consistent with first-order reaction kinetics: $P_r(t) = 1 - \exp(-t/\tau)$, where $\tau \approx 1.6$ h is the characteristic timescale of release (Fig. 1e). Thus, more than 99% of tiles are released after approximately 8 h, and all tiles are released after about 20 h. Although the depletion agent preserves the monolayer, very rarely, a strong local Brownian fluctuation can expel a tile vertically, so a few point defects can be seen for $t > 72$ h. Therefore, by $t \approx 48$ h, many in-plane interparticle collisions have occurred but there are very few defects, so we consider the monolayer to have effectively equilibrated. Despite ongoing Brownian excitations, $\phi_A \approx 0.78$ is large enough that different superstructures of kite–dart motifs, although transiently distorted as tiles explore accessible microstates, remain intact long after release over large length scales (Extended Data Fig. 2) and the confined P2 quasi-crystal of mobile tiles does not melt into a disordered liquid-like state.

Brownian fluctuations of mobile P2 tiles could alter the degree of quasi-crystalline order compared to static tiles, so we take Fourier transforms of optical micrographs to produce the equivalent of light-scattering intensity patterns (see Methods for details). The nearly perfect quasi-crystalline order before release (Fig. 2a) yields Bragg-like peaks at low scattering wavenumbers q, indicative of ordering of larger superstructures of tiles, and ten narrow rays that extend from the centre towards high q, indicative of the high degree of alignment of the edges of all P2 tiles along only specific axes in the plane (Fig. 2b). Applying

[1]Department of Chemistry and Biochemistry, University of California-Los Angeles, Los Angeles, CA, USA. [2]Department of Materials Science and Engineering, University of California-Los Angeles, Los Angeles, CA, USA. [3]Department of Physics and Astronomy, University of California-Los Angeles, Los Angeles, CA, USA. *e-mail: mason@chem.ucla.edu

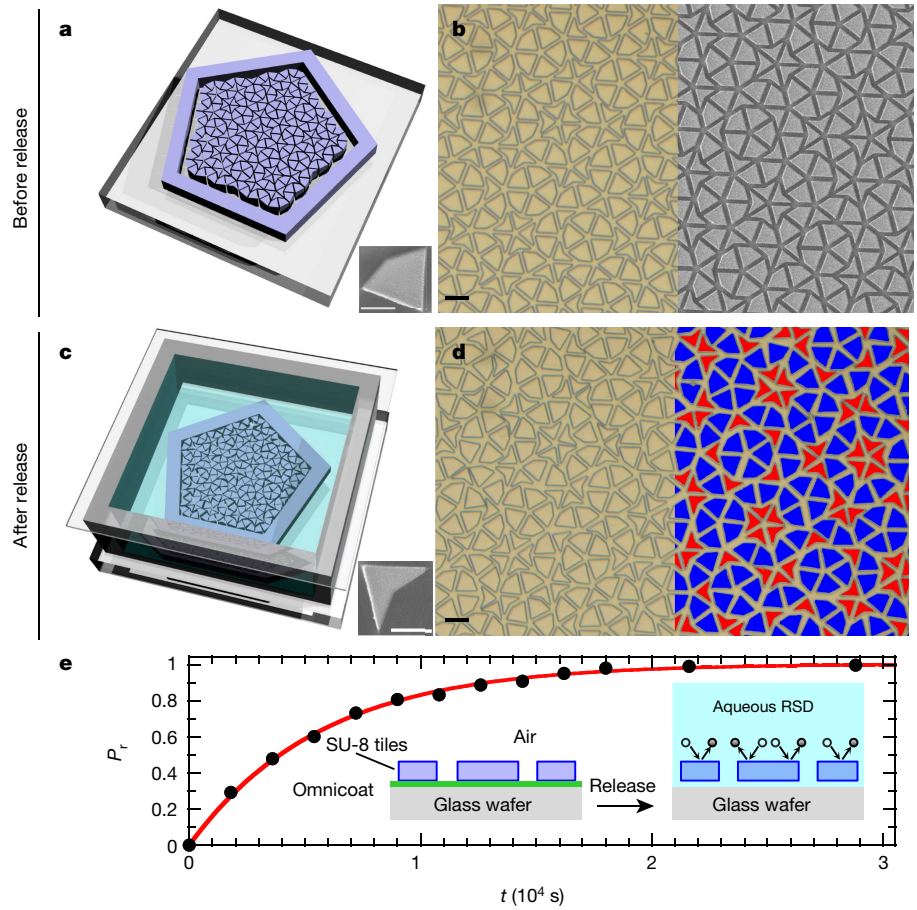


Fig. 1 | Creating a fluctuating Brownian quasi-crystal of mobile Penrose kite and dart tiles in a confined monolayer. a, After designing and fabricating a mask that contains the desired arrangement and area fraction ϕ_A of tiles, ultraviolet stepper lithography is used to print cross-linked polymeric Penrose tile particles in a P2 quasi-crystalline pattern by cross-linking negative SU-8 photoresist at a high ϕ_A inside a pentagonal boundary with an inner edge length of 4.5 mm. Development removes unexposed SU-8 between the tiles, which are attached rigidly to a 10-nm-thick release layer of water-soluble Omnicoat on a glass wafer, thereby enabling observation from below using an inverted bright-field transmission optical microscope. Inset, scanning electron microscope (SEM) image of a kite particle; scale bar, 5 μm. All tiles are 2 μm thick. **b**, Micrographs of pre-configured kite and dart tiles in an ideal quasi-crystalline pattern at $\phi_A \approx 0.78$, after development. Left, optical microscope image; right, SEM image; scale bar, 10 μm. **c**, Solid elastomeric PDMS walls are fabricated to enclose the printed pattern and are attached to the glass. An aqueous RSD is loaded into this PDMS well and a glass coverslip is placed on top to inhibit evaporation. The basic nature of this RSD dissolves the release layer, the surfactant (anionic dodecyl sulfate) adsorbs onto the released tiles and prevents aggregation through screened-charge repulsions, and the nanoparticle dispersion of polystyrene spheres (40-nm diameter, sulfate-stabilized) produces a roughness-controlled depletion attraction between the faces of the tiles and the glass substrate that prevents the released particles from leaving the monolayer, yet keeps their in-plane interactions nearly hard. Inset, SEM image of a dart particle; scale bar, 5 μm. **d**, Optical micrograph taken 48 h after adding the RSD. The right half of the image has been colour-coded (blue, kites; red, darts) using post-acquisition digital analysis; scale bar, 10 μm. **e**, Released fraction of mobile particles P_r as a function of time t. The solid red line is a fit to the data with the function $1 - \exp(-t/\tau)$, assuming first-order reaction kinetics, which yields a release time constant of $\tau = 5{,}680$ s. Inset, schematic side views before and after release. The release layer (green) is dissolved, and the released tiles exhibit Brownian fluctuations in a fully submerged monolayer just above the glass surface, which is negatively charged to ensure that a lubricating layer of water is maintained between the tiles and the glass.

terminology originally developed for systems of anisotropic molecules rather than colloidal tiles, the narrowness of these rays at high q implies that the lithographically pre-assembled P2 tiles have a high degree of long-range molecular orientational order, where here 'molecular orientations' are understood to mean 'tile orientations'. Expanded views at lower q and very low q (Fig. 2c, d) reveal sets of ten Bragg peaks, demonstrating long-range pentagonal quasi-crystalline ordering of motifs and superstructures of motifs.

After release and equilibration, at any given instant the mobile kite and dart tiles are no longer in an ideal P2 tiling (Fig. 2e). As a result, Brownian fluctuations cause the ten rays at high q to broaden azimuthally (Fig. 2f). At intermediate q, Bragg peaks have disappeared and instead ten-fold modulations in ring-like intensity patterns are seen (Fig. 2g), which are reminiscent of the six-fold modulations seen in hexatic liquid-crystal systems[26,27]. Interestingly, large superstructures of motifs retain considerable spatial and orientational order: peaks, although broadened, are still observed at the very lowest q (Fig. 2h). Because entropic–thermal line broadening is very evident at high q, we fit the azimuthal intensity $I(\psi)$ of a ray using a Gaussian function (Fig. 2i; see Methods for details). The peak I of this Gaussian is reduced by a factor of roughly seven and the width ψ_σ is increased by a factor of about two relative to the pre-release intensity profile. At intermediate q, we fit $I(\psi)$ before release with a double Gaussian function to capture the Bragg-like peak; we fit $I(\psi)$ after release using a single Gaussian function (Fig. 2j; see Methods for details). This disappearance of Bragg peaks, yet preservation of a ten-fold modulation in ψ at intermediate q long after release indicates that motifs of mobile tiles are no longer exactly spatially ordered, but that the majority of motifs do preserve a substantial degree of orientational order. Similar broadening effects after release are also evident in Fourier transforms of digitally modified images showing only kite tiles (Extended Data Fig. 3) or only dart tiles (Extended Data Fig. 4). These Fourier transforms long after release provide evidence of the existence of a new phase of matter, which we call a 'pentatic liquid quasi-crystal'.

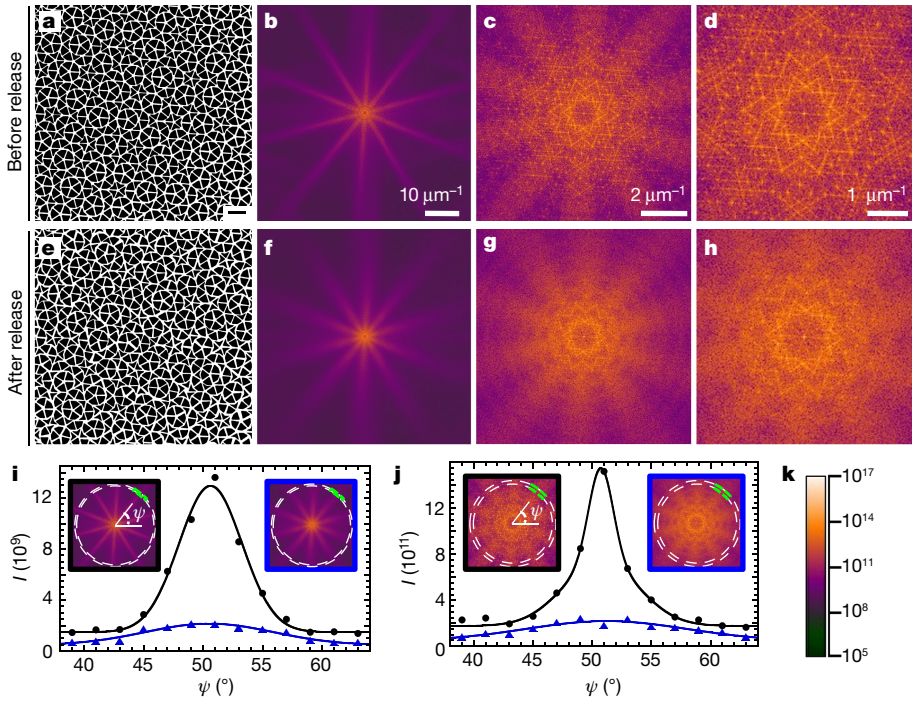


Fig. 2 | Entropic restructuring of ordered Penrose kite and dart tiles into a fluctuating liquid quasi-crystal monolayer after release. **a**, Optical micrograph of Penrose quasi-crystal tiles before release (interiors of tiles are filled black to enhance contrast); scale bar, 20 μm. **b**, Effective scattering pattern provided by the Fourier transform intensity of **a**, showing ten rays extending from the centre to high scattering wavenumbers q. (q is a radial distance measured outward from the centre of the Fourier transform, where $q = 0$; the white scale bar defines the magnitude of q in the Fourier transform.) **c**, Central region of **b**, magnified by a factor of about six, revealing Bragg peaks at intermediate and low q. **d**, Central region of **c**, magnified by a factor of about two, revealing Bragg peaks at very low q associated with superstructural ordering of motifs of tiles over large length scales. **e**, Optical micrograph of fluctuating Penrose quasi-crystal tiles 48 h after release (tile interiors filled black). Entropic Brownian fluctuations destroy the ideal quasi-crystalline order; the tiles no longer have perfect positions and orientations on a quasi-crystalline lattice, but the overall morphology of the quasi-crystal is preserved. **f**, Average Fourier transform intensities of three different micrographs, taken 46 h, 48 h (**e**) and 50 h after release. **g**, **h**, Close-ups of **f**, using the same magnification factors as for **c** and **d**, respectively. All but the superstructural peaks at the very lowest q have become smeared out, leaving only ten-fold azimuthal intensity modulations at intermediate and high q. **i**, High-q azimuthal line shapes $I(\psi)$ of emanating rays before release (black circles, left inset) and after release (blue triangles, right inset). In the insets, the azimuthal angle ψ is defined between white solid lines and the green dashed lines define the annular sectors, inside white dashed lines at high q, used to calculate $I(\psi)$ over the range of ψ displayed. Solid lines in the main panel are fits using Gaussian functions (see text). **j**, Intermediate-q $I(\psi)$ (insets as in **i**) before release (black circles) and after release (blue triangles). Black line, fit using a double Gaussian; blue line, fit using a Gaussian (see text). **k**, Intensity scale for all Fourier transforms shown.

Equilibrium Brownian forces cause motifs—such as the pentagonal flower composed of 15 kite tiles and five dart tiles (Fig. 3a, top; Supplementary Video 2) and the pentagonal wheel composed of ten kite tiles and five dart tiles (Fig. 3a, bottom; Supplementary Videos 3, 4)—to fluctuate and distort randomly, breaking chiral symmetry locally without melting[28]. Within the fluctuating P2 system, we observe that five kites in pentagonal star-like motifs (PSKMs) can rotate collectively, and the entire motif makes rotational transitions between different preferred angles, defined by corrugations in proximate boundary tiles (Fig. 3b, top; Supplementary Video 5). We measure the heterogeneous dynamics[29] of these collective rotational fluctuations by digitally tracking the rotational angle α and the trajectory of the centroid of one kite in the PSKM over time (Fig. 3b, bottom; Fig. 3c). By contrast, five darts in pentagonal star-like motifs (PSDMs), which have much more corrugated exteriors, fluctuate and distort, which leads to bounded Brownian motion that is anisotropic and reflects the local symmetry of the fluctuating quasi-crystal (Methods, Supplementary Video 4, Extended Data Fig. 5). Although all motifs distort, and PSKMs also rotate, larger-scale superstructures made up of multiple motifs appear to preserve long-range orientational order. To quantify this, we create a motif superstructural orientational pair-correlation function $g_{\mathrm{MSO\text{-}PCF}}$ that depends on the separation r between the centres of two similar superstructures of motifs (Supplementary Methods). For instance, we define a superstructure of five PSKMs in a nearly regular pentagonal configuration that surrounds a central PSKM (Fig. 3d); superstructural bond angles (Fig. 3d, defined by dashed arrows) are then correlated in a manner analogous to bond orientational pair-correlation functions of individual particles. Before release $g_{\mathrm{MSO\text{-}PCF}}(r) = 1$, whereas after release $g_{\mathrm{MSO\text{-}PCF}}(r)$ decays exponentially to a plateau that is just below unity, indicating strong preservation of relative superstructural bond angles over long distances (Fig. 3e). Therefore, motif superstructural orientational order is preserved over long times and distances even as distortions and rotations of smaller constituent motifs occur. Moreover, because the P2 quasi-crystal of mobile tiles exhibits solid-like behaviour, similar to simpler dispersed systems of colloidal hard spheres at high densities[30], we observe entropically generated sound wavelets that scatter locally and are damped (Supplementary Video 6).

We cause dense P2 quasi-crystals to melt by removing a confining wall. Tiles diffuse gradually into the available empty space over time, thereby creating a gradient in ϕ_{A} (Fig. 4a–c, Supplementary Video 7). Digitally colour-coded PSKMs (blue) and PSDMs (red) progressively melt. We measure ϕ_{A} as a function of distance d at different t (Fig. 4d), and the area fractions of intact PSKMs $\phi_{\mathrm{A,KM}}$ (Fig. 4e) and PSDMs $\phi_{\mathrm{A,DM}}$ (Extended Data Fig. 6). Fermi functions, consistent with diffusive melting, fit the measured $\phi_{\mathrm{A}}(d)$, $\phi_{\mathrm{A,KM}}(d)$ and $\phi_{\mathrm{A,DM}}$ profiles (solid lines in Fig. 4d, e, Extended Data Fig. 6; see Methods for details and Extended Data Table 1 for fit parameters). Fourier transforms of the same region before, during and after the passage of the melting front (Fig. 4f–h) reveal the progressive disappearance of spatial and orientational order. Peaks at very low q, which were evident before melting

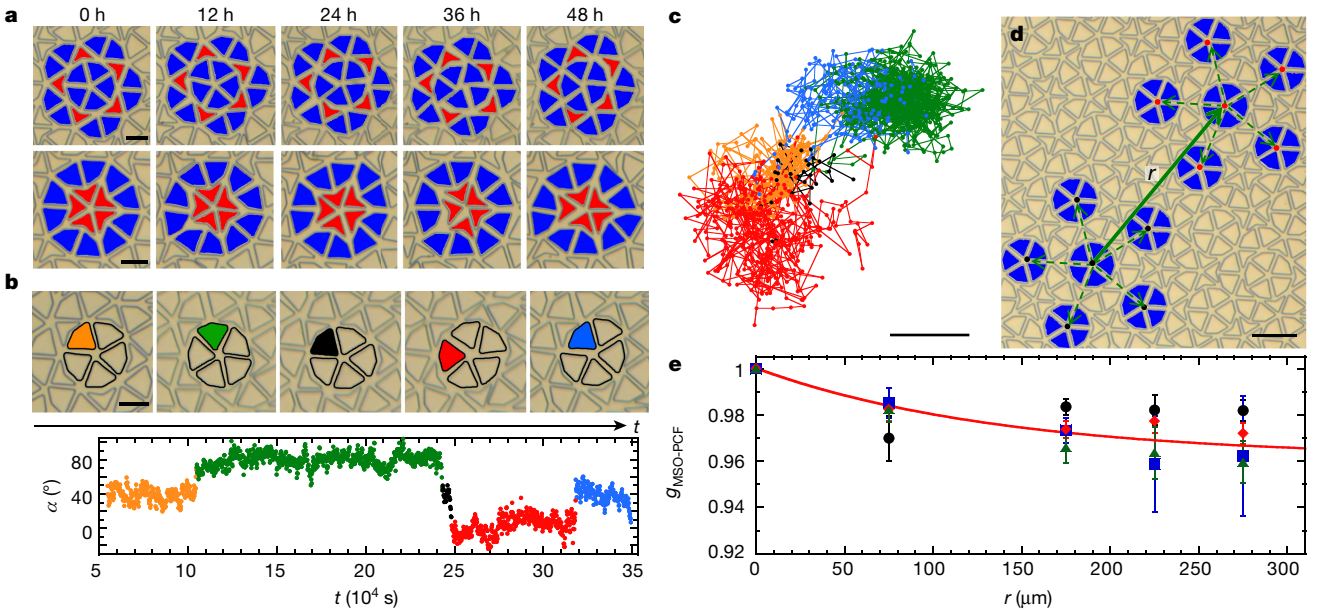


Fig. 3 | Motif dynamics and superstructural orientational pair-correlation function. a, Optical micrographs of fluctuating pentagonal motifs of Penrose kite and dart tiles before release (0 h), and 12 h, 24 h, 36 h and 48 h after release. Top, flower motif consisting of 15 kite tiles and five dart tiles; bottom, wheel motif consisting of ten kite tiles and five dart tiles. Scale bar; 10 μm. Entropic fluctuations lead to local chiral-symmetry breaking of motifs. **b**, Collective rotational fluctuations of a central pentagonal star of kites (black outlines) within a flower. One particular kite is coloured to show the collective rotation of the entire motif over time t (top); its rotational angle α versus t (bottom) exhibits hopping behaviour between preferred angles (plateaus separated by about 36°) that match corrugations of the surrounding tiles. Scale bar, 10 μm. **c**, Trajectory of the centroid of the coloured kite in **b**, which displays local heterogeneous dynamics. Scale bar, 3 μm. **d**, One pentagonal superstructure of PSKMs (kites filled blue, black dot at centre) is separated by a centre-to-centre distance r from a second pentagonal superstructure of PSKMs (kites filled blue, red dot at centre). Solid green arrow, vector between superstructure centres; dashed green arrows, vectors between superstructure centres and PSKM centres. Scale bar, 20 μm. **e**, Motif superstructural orientational pair-correlation function $g_{\text{MSO-PCF}}$ as a function of r at different times t after release: black circles, 24 h; blue squares, 48 h; red diamonds, 72 h; green triangles, 96 h. Before release, $g_{\text{MSO-PCF}} = 1$ for all r. After release, random Brownian excitations of the system cause slight fluctuations in the orientations of superstructures of motifs. Error bars, s.d. for $r > 0$, $g_{\text{MSO-PCF}} = 1$ exactly by definition for $r = 0$); statistical uncertainties largely overlap for $t \geq 48$ h. Solid red line, fit to the average of the data at 48 h and 72 h using exponential decay of $g_{\text{MSO-PCF}}$ from 1 to a plateau of $g_{\text{MSO-PCF}} = 0.96$ at large r. The high and persistent value of $g_{\text{MSO-PCF}}$ near but below unity for large r over long times indicates that long-range orientational order of superstructures of motifs persists in equilibrium after release, even as perfect tile–tile spatial order is destroyed over long distances as a consequence of the entropic Brownian fluctuations of the tiles.

(Fig. 4f), disappear, and only isotropic rings, indicative of liquid-like disorder, appear in the Fourier transforms immediately after melting (Fig. 4h). To determine the value of ϕ_A associated with melting, we eliminate d and plot $\phi_{A,KM}$ versus ϕ_A at different t (Fig. 4i). For longer t, towards equilibrium, the measured $\phi_{A,KM}(\phi_A)$ curves almost overlap; so, by considering $\phi_{A,KM} \rightarrow 0$, we find that the P2 quasi-crystal effectively melts at $\phi_{A,melt} \approx 0.65 \pm 0.02$ (here and elsewhere, the values quoted are mean ± s.d.). This melting and the diffusion of individual tiles into the empty space provides additional empirical evidence that any residual in-plane depletion attractions between tiles are weak compared to thermal energy k_BT, where k_B is Boltzmann's constant and T is the temperature (Supplementary Information).

The litho-PAM method that we have developed to make a fluctuating P2 quasi-crystal can readily be generalized to provide diverse, complex and essentially defect-free experimental soft-matter systems that are suitable for testing predictions of theories and simulations, including fluctuating dynamics and kinetics. Specifically, litho-PAMs can be used to reveal and measure localized and anisotropic translational and rotational dynamics associated with different particle shapes or local groups of shapes, some of which can move collectively, over a wide range of densities and length scales. These observations could motivate improvements in simulations, which sometimes do not report results in units of real time that can be compared directly with experiments. Detailed statistical analyses of measured trajectories of tile collisions at different relative positions and orientations in experiments could potentially be compared to simulations to deduce and quantify interactions between neighbouring tiles, such as hydrodynamic interactions and site-specific attractive or repulsive interactions.

Extending theories of multi-body systems of single colloidal shapes to multiscale systems containing two or more different tile shapes that have hard or nearly hard in-plane interactions, such as the system that we have demonstrated, represents an exciting challenge. For such fluctuating monolayers of hard tiles, the free energy is entirely entropic in origin, which is considerably different from the free energy of strongly bonded configurations of atoms in classic solid-state quasi-crystals[3] (see Methods section 'Additional discussion'). Ultimately, theories of entropy-dominated fluctuating monolayers of complex shapes could be used to predict the phase behaviour, dynamic heterogeneity and melting of the soft-matter P2 quasi-crystal and other multiscale structures made using litho-PAMs.

Although we have demonstrated one viable route for creating litho-PAMs using optical stepper lithography and a specifically formulated RSD that creates anisotropic roughness-controlled depletion attractions, the general concept of fluctuating litho-PAMs is not limited to these specific methods of fabrication, these materials or this particular type of interaction. Other methods of fabrication, such as forms of nanolithography (see Supplementary Discussion), could also be used. The tiles do not need to be slab-like, but instead could have more complex, out-of-plane shapes, and their overall lateral dimensions could be much smaller. Beyond cross-linked polymeric photoresist, the tiles could alternatively be composed of inorganic materials or metals. Likewise, other release materials and agents could be used. Although anisotropic roughness-controlled depletion attractions provide a convenient combination of strong out-of-plane 'lubricated' attractions between the faces of the tiles and the substrate (so the tiles do not stick rigidly to the substrate but instead remain attractively bound to

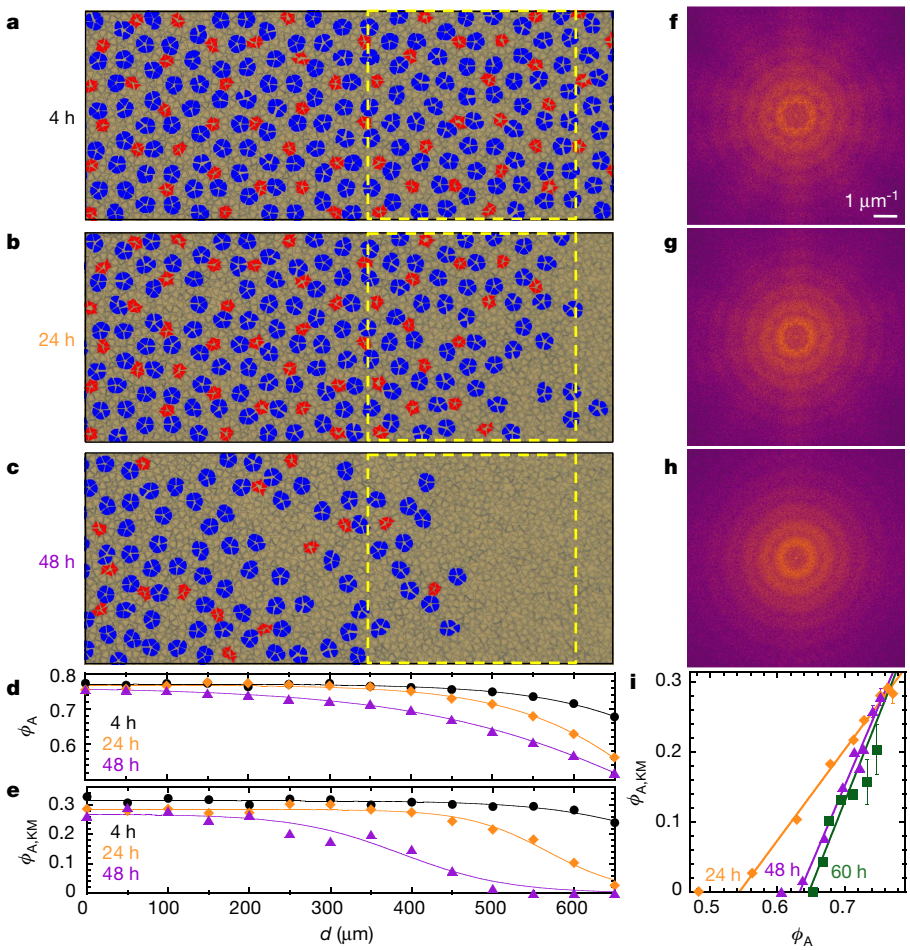


Fig. 4 | Entropic melting dynamics of an unconfined fluctuating quasi-crystal. a–c, Colour-coded optical micrographs of Penrose kite and dart tiles that melt diffusively into an unconfined open region beyond the right edge of the micrograph, at different times after release: **a**, 4 h; **b**, 24 h; **c**, 48 h. Blue, PSKMs; red, PSDMs. **d**, Area fraction ϕ_A as a function of distance d measured from the left of the micrographs in **a**–**c**, at different times after release: black circles, 4 h; orange diamonds, 24 h; purple triangles, 48 h. Lines are fits using an empirical Fermi-like function (Methods). **e**, Area fraction of unmelted PSKMs $\phi_{A,KM}$ as a function of d at different times after release (symbols and lines as in **d**). **f**, Fourier transform of the image in **a** inside the yellow dashed square. Ten-fold azimuthal intensity modulations are evident. **g**, Fourier transform of the image in **b** inside the yellow dashed square. The tenfold azimuthal intensity modulations are less pronounced. **h**, Fourier transform of the image in **c** inside the yellow dashed square. The tenfold azimuthal intensity modulations are no longer visible, and the ring-like pattern indicates that the tiles in the region are largely disordered, as in a glass or liquid. **i**, To determine the melting area fraction $\phi_{A,melt}$ of the P2 quasi-crystal, we plot $\phi_{A,KM}$ as a function of ϕ_A at various times after release: green circles, 24 h; orange diamonds, 48 h; purple triangles, 60 h. Error bars, s.d. At the two longest times, we find nearly the same melting behaviour, yielding an intercept with the horizontal axis of $\phi_{A,melt} \approx 0.65$. Lines, linear fits (see Methods).

it while maintaining a layer of liquid in between) and only very weak residual in-plane attractions, their use is not an inherent aspect of litho-PAMs. Other types of anisotropic attraction that can provide lubricated attractions between the tiles and the substrate, but do not cause strong in-plane attractions compared to the thermal energy that would lead to the aggregation of tiles, could also be used. The sizes and shapes of the boundary walls can also be controlled lithographically, opening the door to visualizing entropic and steric effects in specially confined systems of shape-designed Brownian particles.

As the desired level of complexity in multiscale materials increases, some limitations of previous self- and directed-assembly methods have become apparent. The litho-PAM method described here represents an alternative approach that can provide access to a wide range of two-dimensional multiscale materials composed of differently shaped mobile tiles, including fluctuating Brownian systems that have unusual symmetries and hierarchical structures. Thus, litho-PAMs can be used to create and study new equilibrium phases of such systems, including the fluctuating P2 quasi-crystal that we have demonstrated here. These new phases could potentially display unusual spatio-temporal dynamics at different length scales, such as the transient collective hopping motion of the PSKMs in our Brownian P2 quasi-crystal. Using litho-PAMs to systematically control the degree of coupling between different types of motif and the rest of the system is likely to lead to a broader understanding of heterogeneous dynamics in complex multiscale Brownian systems, beyond disordered glassy systems. Moreover, the litho-PAM method opens up possibilities for exploring and directly visualizing the evolution of multiscale systems that have pre-assembled out-of-equilibrium initial states, which would otherwise be very difficult to create (see Methods section 'Additional discussion'). Although our demonstration of litho-PAMs has focused on a quasi-crystal system consisting of only two tile shapes, a much larger number of different tile shapes could be simultaneously fabricated and pre-configured using litho-PAMs, yielding systems with even higher levels of multiscale morphological complexity, including sub-particle features in tile shapes, local multi-tile motifs and intricate and diverse superstructures.

Online content

Any Methods, including any statements of data availability and Nature Research reporting summaries, along with any additional references and Source Data files, are available in the online version of the paper at https://doi.org/10.1038/s41586-018-0464-9.

Received: 30 November 2017; Accepted: 13 July 2018;
Published online 29 August 2018.

1. Penrose, R. The role of aesthetics in pure and applied mathematical research. *Bull. Inst. Math. Appl.* **10**, 266–271 (1974).
2. Penrose, R. Pentaplexity. A class of non-periodic tilings of the plane. *Math. Intell.* **2**, 32–37 (1979).
3. Bursill, L. A. & Lin, P. J. Penrose tiling observed in a quasi-crystal. *Nature* **316**, 50–51 (1985).
4. Levine, D. & Steinhardt, P. J. Quasi-crystals: a new class of ordered structures. *Phys. Rev. Lett.* **53**, 2477–2480 (1984).
5. Zeng, X. et al. Supramolecular dendritic liquid quasicrystals. *Nature* **428**, 157–160 (2004).
6. Hayashida, K., Dotera, T., Takano, A. & Matsushita, Y. Polymeric quasicrystal: mesoscopic quasicrystalline tiling in ABC star polymers. *Phys. Rev. Lett.* **98**, 195502 (2007).
7. Talapin, D. V. et al. Quasicrystalline order in self-assembled binary nanoparticle superlattices. *Nature* **461**, 964–967 (2009).
8. Fischer, S. et al. Colloidal quasicrystals with 12-fold and 18-fold diffraction symmetry. *Proc. Natl Acad. Sci. USA* **108**, 1810–1814 (2011).
9. Hoover, M. D. et al. A method for producing non-spherical monodisperse particles using integrated circuit fabrication techniques. *J. Aerosol Sci.* **21**, 569–575 (1990).
10. Hernandez, C. J. & Mason, T. G. Colloidal alphabet soup: monodisperse dispersions of shape-designed lithoparticles. *J. Phys. Chem. C* **111**, 4477–4480 (2007).
11. Yang, S.-M., Kim, S.-H., Lim, J.-M. & Yi, G.-R. Synthesis and assembly of structured colloidal particles. *J. Mater. Chem.* **18**, 2177–2190 (2008).
12. Zeng, C., Chen, Y., Kirschbaum, K., Lambright, K. J. & Jin, R. Emergence of hierarchical structural complexities in nanoparticles and their assembly. *Science* **354**, 1580–1584 (2016).
13. Grier, D. G. A revolution in optical manipulation. *Nature* **424**, 810–816 (2003).
14. Bae, W.-G. et al. Scalable multiscale patterned structures inspired by nature: the role of hierarchy. *Adv. Mater.* **26**, 675–700 (2014).
15. Roichman, Y. & Grier, D. G. Holographic assembly of quasicrystalline photonic heterostructures. *Opt. Express* **13**, 5434–5439 (2005).
16. Mikhael, J., Roth, J., Helden, L. & Bechinger, C. Archimedean-like tiling on decagonal quasicrystalline surfaces. *Nature* **454**, 501–504 (2008).
17. Wen, W., Zhang, L. & Sheng, P. Planar magnetic colloidal crystals. *Phys. Rev. Lett.* **85**, 5464–5467 (2000).
18. Boles, M. A., Engel, M. & Talapin, D. V. Self-assembly of colloidal nanocrystals: from intricate structures to functional materials. *Chem. Rev.* **116**, 11220–11289 (2016).
19. Niederberger, M. Multiscale nanoparticle assembly: from particulate precise manufacturing to colloidal processing. *Adv. Funct. Mater.* **27**, 1703647 (2017).
20. Grzelczak, M., Vermant, J., Furst, E. M. & Liz-Marzán, L. M. Directed self-assembly of nanoparticles. *ACS Nano* **4**, 3591–3605 (2010).
21. Campbell, M., Sharp, D. N., Harrison, M. T., Denning, R. G. & Turberfield, A. J. Fabrication of photonic crystals for the visible spectrum by holographic lithography. *Nature* **404**, 53–56 (2000).
22. Ullal, C. K. et al. Photonic crystals through holographic lithography: simple cubic, diamond-like, and gyroid-like structures. *Appl. Phys. Lett.* **84**, 5434–5436 (2004).
23. Kawata, S., Sun, H.-B., Tanaka, T. & Takada, K. Finer features for functional microdevices. *Nature* **412**, 697–698 (2001).
24. Zhao, K. & Mason, T. G. Directing colloidal self-assembly through roughness-controlled depletion attractions. *Phys. Rev. Lett.* **99**, 268301 (2007).
25. Zhao, K. & Mason, T. G. Frustrated rotator crystals and glasses of Brownian pentagons. *Phys. Rev. Lett.* **103**, 208302 (2009).
26. Nelson, D. R., Rubinstein, M. & Spaepen, F. Order in two-dimensional binary random arrays. *Philos. Mag. A* **46**, 105–126 (1982).
27. Nelson, D. R. *Defects and Geometry in Condensed Matter Physics* Ch. 2 (Cambridge Univ. Press, New York, 2002).
28. Zhao, K., Bruinsma, R. & Mason, T. G. Local chiral symmetry breaking in triatic liquid crystals. *Nat. Commun.* **3**, 801 (2012).
29. Weeks, E. R., Crocker, J. C., Levitt, A. C., Schofield, A. & Weitz, D. A. Three-dimensional direct imaging of structural relaxation near the colloidal glass transition. *Science* **287**, 627–631 (2000).
30. Chen, K. et al. Low-frequency vibrations of soft colloidal glasses. *Phys. Rev. Lett.* **105**, 025501 (2010).

Acknowledgements We thank C. M. Knobler for discussions and UCLA for financial support. We acknowledge the use of the Integrated Systems Nanofabrication Cleanroom at the California NanoSystems Institute at UCLA.

Author contributions T.G.M. conceived the experimental approach and supervised the project. P.-Y.W. and T.G.M. designed the experiments. P.-Y.W. performed the experiments. P.-Y.W. and T.G.M. analysed the data. P.-Y.W. and T.G.M. wrote the manuscript.

Competing interests There is a pending and unlicensed US provisional patent application assigned to and filed by UCLA relating to this work.

Additional information
Extended data is available for this paper at https://doi.org/10.1038/s41586-018-0464-9.
Supplementary information is available for this paper at https://doi.org/10.1038/s41586-018-0464-9.
Reprints and permissions information is available at http://www.nature.com/reprints.
Correspondence and requests for materials should be addressed to T.G.M.
Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

METHODS

Mask design. Microscale kite-shaped and dart-shaped prismatic tiles are fabricated out of epoxy-based negative photoresist SU-8 (Microchem), and these tiles are spatially and orientationally organized before release using the following top-down photolithography process. A quartz-chrome lithographic mask (150 mm × 150 mm) containing a pattern of kite and dart tiles, based on Penrose's P2 quasi-crystalline pattern, is designed at a desired tile area fraction $\phi_{A,des}$. Each kite is a convex quadrilateral, with two adjacent sides of designed shorter length C_{des}, two adjacent sides of designed longer length D_{des}, a 144° interior angle between the two shorter sides and other interior angles of 72° (Fig. 1a, inset). The ratio of the two side lengths D_{des}/C_{des} is the golden ratio (1.618...). Each dart is a concave quadrilateral with side lengths C'_{des} and D'_{des} that are nearly the same as C_{des} and D_{des}, and interior angles of 216° (concave vertex), 72° (convex vertex opposite the concave vertex) and two of 36° (convex vertices at the two sharp tips) (Fig. 1c, inset). A P2 quasi-crystalline pattern containing about 10^6 kite and dart tiles is digitally designed using lithographic layout software (L-Edit, Tanner Research). An idealized P2 pattern is generated using thin lines, and these lines are subsequently thickened so that the tiles outside these lines have the desired area fraction, in our case $\phi_{A,des} \approx 0.78$. Then, we determine the coordinates of the vertices of the kite and dart tiles just inside these thickened lines. We use $C_{des} = 31.3\,\mu m$, $D_{des} = 50.7\,\mu m$, $C'_{des} = 29.4\,\mu m$ and $D'_{des} = 47.6\,\mu m$ in the quartz-chrome masks; after five-times-reduction stepper optical lithographic printing, these mask dimensions lead to ideal edge lengths of $C \approx 6.2\,\mu m$ and $D \approx 10.1\,\mu m$ for the prismatic SU-8 kite particles, and of $C' \approx 5.8\,\mu m$ and $D' \approx 9.5\,\mu m$ for the prismatic SU-8 dart particles. The overall lateral dimensions of the SU-8 tiles were chosen to ensure that diffraction during the UV printing process did not cause the tiles to become irreversibly bonded together via unwanted cross-linking in the spaces between them when printed at high densities. The separation between the printed SU-8 tiles at $\phi_A \approx 0.78$ is uniformly about 1.2 μm, which is about four times the minimum feature size that can be printed by the particular UV stepper that we used. Thus, although our pre-assembled pattern of kite and dart tiles is inspired by the classic P2 construction that involves infinitesimally thin lines, it is distinguishably different from it because our variation enables us to set $\phi_{A,des}$ and to determine the vertices of all constituent tiles in a well-defined manner. To confine the mobile kite and dart particles and to maintain a fixed ϕ_A after release, we designed millimetre-scale pentagonal-shaped confining walls. Each confining wall has a lateral thickness of about 100 μm and an inner edge length of about 2 mm, making contact with the outermost confined Penrose tiles. This confining boundary is also made of SU-8 photoresist and is printed together with the particles in the same exposure step.

Lithographic production. To observe the kite and dart particles before, during and after release, we use a transparent glass wafer as a substrate and record in situ time-lapse images using bright-field transmission microscopy. A clean glass wafer (100-mm diameter, 500 μm thick, borosilicate) is pre-baked on a hot plate at 200 °C for 3 min to remove adsorbed moisture from its surface and then cooled to 25 °C. Although other smooth substrates can be used (such as silicon wafers), we chose a glass wafer to facilitate visualization of the monolayer using an inverted optical microscope and because glass becomes negatively charged in the aqueous basic RSD used in a subsequent step to release the tiles (consequently, anionically stabilized tiles will not stick to the glass surface). Onto this glass wafer we spin-coat (Headway Research, PWM32 Spinner; 500 r.p.m. for 5 s at an acceleration of 100 r.p.m. s^{-1}; 3,000 r.p.m. for 30 s at an acceleration of 300 r.p.m. s^{-1}) a release layer of Omnicoat (Microchem), yielding an optically transparent thin layer with a thickness of about 10 nm after baking at 200 °C for 1 min and then cooling to 25 °C. Subsequently, we spin-coat a layer of SU-8-2002 (Microchem) negative photoresist on top of the Omnicoat layer and then bake at 95 °C for 90 s, yielding an optically transparent, un-cross-linked solid layer of SU-8 that is approximately 2 μm thick after cooling to 25 °C. This coated wafer is exposed to patterned UV light (365 nm, typical energy dose of 180 mJ cm^{-2}) through the designed photomask (Digidat, 150 mm × 150 mm × 6.25 mm, chrome on quartz) in a lithographic stepper (ASML, PAS 5500/200, five times reduction, Hg i-line).

Our printing routine on the stepper yields 14 identical dies containing kite and dart particles within confining boundaries; these dies are evenly distributed on the wafer and separated by about 18 mm. The patterned UV light produced by the stepper triggers cross-linking of the oligomeric epoxy SU-8 photoresist, and the wafer is baked post-exposure at 95 °C for 75 s to enhance the rate of cross-linking reactions. After cooling to 25 °C, the post-exposure-baked wafer is submerged in organic SU-8 developer (1-methoxy-2-propyl acetate, Microchem) for 4 min to remove the un-cross-linked SU-8, rinsed with isopropyl alcohol to wash away any residual developer on the wafer and dried using a nitrogen gas stream. The result is a set of discrete prismatic kite and dart particles in a Penrose P2 pattern, bound to the Omnicoat release material on the wafer, yet entirely disconnected from each other, despite the high area fraction. To achieve this, we adjusted the energy dose and post-bake duration for SU-8 as described above, optimizing this lithography process for transparent glass wafers. We designed the release process so that the thin Omnicoat release layer dissolves in an aqueous solvent rather than the organic solvent (cyclopentantone) or developer for the SU-8 resist layer, so unexposed resist can be completely removed through development and the resulting SU-8 particles can be released independently. We kept the thickness of the layer of release material to a minimum (about 10 nm) that still provides full release of all tiles yet minimizes the amount of dissolved release material in the RSD and thereby reduces associated fluid flows due to concentration gradients. Transmission optical microscopy through the glass wafer in an inverted imaging configuration provides better-quality images of the SU-8 particles compared to reflection optical microscopy of the particles through a thicker layer of RSD above an absorbing and reflecting silicon wafer. The actual ϕ_A of printed tiles, which have the same P2 arrangement as defined in the mask, can be adjusted over a limited range near $\phi_{A,des}$ by tuning the energy dose of the exposure through the stepper and by adjusting the post-exposure baking conditions.

Releasing the tiles to form a fluctuating monolayer. To obtain large areas of stable fluctuating particles in a monolayer, controlling their release is essential. If the release occurs too rapidly, then concentration gradients of the release material can drive discrete particles out of the monolayer, leading to undesirable vacancies and defects in the overall structure. The composition of the RSD is therefore important, so including a stabilizing agent, to prevent irreversible aggregation of the tiles, as well as a depletion agent, to overcome out-of-plane entropic excitations and keep tiles in the monolayer, is typically necessary.

After development of the SU-8 photoresist, but before release, around a given die on the dry glass wafer we build a solid square-frame enclosure of four elastomeric PDMS walls. The main purpose of these PDMS walls is to confine the RSD that is added later to release the SU-8 particles. We first make a solid disk-like layer of PDMS elastomer that is about 3 mm thick and 5.5 cm in diameter in a small Petri dish, then cut this into a square frame (inner edge length of 1.3 cm, lateral wall thickness of 0.2 cm), remove the square frame from the Petri dish and attach it to the glass wafer surrounding the die of SU-8 particles. To make the disk-like layer of PDMS, we mix a PDMS elastomer and its curing agent (Sylgard 184, weight ratio of 10:1 of elastomer:curing agent) on a clean Petri dish, gradually de-gas the PDMS at 25 °C for 12 h, and then place the Petri dish in a vacuum oven at 80 °C for 2 h to remove any remaining entrapped bubbles and to cure the PDMS. Before attaching the PDMS walls to the glass wafer, we remove the Omnicoat layer outside the die region using a cotton-tipped applicator soaked in release solution. The PDMS walls form a water-impermeable bond with the glass wafer around a given die, and the bonded PDMS square frame and glass wafer effectively form a chamber (a PDMS well) into which the RSD can be loaded. An adequate volume of the RSD is placed in contact with the tiles attached to the release layer, so that components in the RSD are in excess and not consumed. A coverslip is placed on top of the PDMS walls and in contact with the RSD to inhibit evaporation of the RSD and to prevent convective flows during and after the release process. Inside the PDMS walls, the lithographically constructed SU-8 walls confine the particles to a fixed ϕ_A. The size and location of the PDMS walls can be chosen to anchor the SU-8 walls to the glass wafer.

To release the tiles slowly but preserve their pre-assembled organization in the monolayer without introducing strong in-plane attractions between them, we formulated an aqueous RSD that contains: a basic release agent (tetramethylammonium hydroxide, TMAH; Sigma-Aldrich, 1% w/v) that dissolves Omnicoat, a stabilizing agent (sodium dodecyl sulfate, SDS; MP Biomedicals, ultrapure, 10 mM) that prevents irreversible aggregation of released SU-8 particles, and a depletion agent (carboxylate-stabilized, surfactant-free polystyrene spheres, Magsphere, 52-nm diameter, 1% w/v solids) that strongly inhibits the released SU-8 particles from being entropically excited in the direction normal to the surface of the glass and out of the monolayer. We fill the PDMS well completely with this RSD and immediately place a coverslip over it to reduce evaporation, which can lead to undesirable convection. As the SU-8 particles are released, dodecyl sulfate anions adsorb onto their surfaces, providing stabilizing screened-charge electrostatic repulsions and preventing irreversible aggregation of the particles when they collide with each other in the monolayer. The depletion agent induces a roughness-controlled depletion attraction that causes a preferential attraction of a face of each SU-8 particle towards the flat smooth surface of the glass wafer, whereas in-plane attractions between rougher edges of particles are small compared to thermal energy k_BT, where $T \approx 298$ K is the temperature. We optimized the concentrations of SDS, the diameter and volume fraction of the depletion agent, the thickness of the SU-8 particles, and the type and thickness of the release layer to keep all of the released particles in the monolayer for an extended period of time (at least several days). Because rapidly dissolving the Omnicoat can drive undesirable fluid flows that cause tiles to lift off vertically out of the monolayer, we adjusted the TMAH concentration so that tiles release gradually. Moreover, the size and volume fraction of the depletion agent were set to ensure that each tile experiences a strong facial depletion attraction with the smooth glass substrate that

is substantially larger than k_BT; however, in-plane depletion attractions with other tiles are much smaller, because the edges of tiles are rougher than the faces (see supporting information in refs [24,25]). The walls of the PDMS well were designed to be tall enough (3 mm high) to keep any residual convection near the coverslip away from the monolayer, thereby enabling time-lapse imaging of the Brownian system of tiles over many days.

High-resolution time-lapse particle imaging by optical bright-field transmission microscopy. After filling and covering the PDMS well, in situ images are taken before, during and after particle release using an inverted bright-field microscope (Nikon Eclipse TE2000) equipped with a 20× objective lens (NIKON CFI60 Plan Achromat, 0.4 numerical aperture), a 10× objective lens (NIKON Plan Achromat, 0.25 numerical aperture) and a Nikon D5000 camera (4,288 pixels × 2,848 pixels, silent mode). The image contrast and resolution are sufficiently large that even for the lower-power 10× objective, the position and orientation of each particle can be readily seen over the entire field of view. Individual images showing a fixed region of Penrose tiles are taken at a rate of one frame every 60 s using a computer-automated camera control system.

Measuring area fractions of Penrose tiles. The interiors of all tiles in an optical micrograph are filled with coloured pixels using an edge-detection routine (paint bucket fill tool) in Photoshop, and a first estimate of the area fraction is determined by counting these coloured pixels and dividing by the total number of pixels. The area fraction of tiles from optical micrographs before release is $\phi_{A,OM} = 0.677 \pm 0.009$. This value is roughly 13% lower than the designed area fraction on the mask of $\phi_{A,des} = 0.78$, owing to optical diffraction during the lithographic printing process and to the dose response of the SU-8 photoresist. To improve this first estimate, we take and analyse a high-resolution SEM image (magnification 1,000×), which does not suffer from optical diffraction, to determine the area fraction of tiles, measured at their upper surfaces (tops), yielding $\phi_{A,SEMtop} = 0.684 \pm 0.003$. The factor $f_{SEMtop-OM} = \phi_{A,SEMtop}/\phi_{A,OM} = 1.01$ is used to correct the uncertainties that result from optical diffraction from the particle edges in the optical micrographs.

Because the SEM images of particles before release provide areas of only the tops of the tiles, they cannot reveal protruding features on the sides of the particles that correspond to edge roughness, a natural consequence of the photolithographic exposure and development processes. However, such edge roughness can affect the steric interactions between adjacent tiles. To measure a tile area fraction that could best be compared to future simulations of tiles with hard interactions, we use side or oblique SEM imaging of tiles that are lifted off the substrate after release to estimate the average edge roughness (about 180 nm). From these SEM measurements, we determine a correction factor for the area fraction as a consequence of edge roughness of $f_{ER} = 1.13$, The measured area fraction ϕ_A is then determined from the filled tiles in the optical micrographs by combing the two correction factors: $\phi_A = f_{SEMtop-OM} f_{ER} \phi_{A,OM}$. Thus, the ϕ_A values that we report include corrections for the dose response of the photoresist, blurring caused by optical diffraction and the roughness of the edges of the tiles.

Fourier transforms of microscope images containing filled kite and dart tiles. Measured images in 24-bit RGB colour are converted to unsigned 8-bit greyscale and then cropped into a square region of 2,048 pixels × 2,048 pixels. To enhance contrast in an image, the interiors of the particles are filled with black and the regions outside the particles are filled with white in Photoshop. We use ImageJ to perform a 2D fast Fourier Transform (FFT) of these black-and-white images, yielding intensity and phase information. We display the FFT intensity using a colour lookup table, to convert the greyscale intensity values to colours. In some cases, we average the FFT intensities of several images taken at different times.

Fitting azimuthally varying intensities of rays and peaks in Fourier transforms. In Fig. 2i, we fit the ψ-dependent intensities (azimuthal line shapes) of rays at high q using a single Gaussian function with a constant background I_0: $I(\psi) = I_0 + I_h \exp[-(\psi - \psi_p)^2/(2\psi_\sigma^2)]$, where I_h is the peak intensity relative to I_0, ψ_p is the mean peak angle and ψ_σ is the standard deviation in angle (the effective azimithal width of a ray). A fit to the data in Fig. 2i before release yields $I_0 = (1.48 \pm 0.13) \times 10^9$, $I_h = (11.5 \pm 0.3) \times 10^9$, $\psi_p = 50.6° \pm 0.1°$ and $\psi_\sigma = 2.6° \pm 0.1°$, with a correlation coefficient of $R^2 = 0.996$. A fit to the data in Fig. 2i after release yields $I_0 = (0.45 \pm 0.09) \times 10^9$, $I_h = (1.70 \pm 0.11) \times 10^9$, $\psi_p = 50.5° \pm 0.3°$ and $\psi_\sigma = 5.4° \pm 0.5°$, with $R^2 = 0.977$. Entropic fluctuations after release therefore cause a noticeable increase in the width and a decrease in the central intensity of the rays at high q.

In Fig. 2j, for the FFT before release, we semi-empirically fit the azimuthal line shape at intermediate q using a double Gaussian function: $I_0 + I_{h1}\exp[-(\psi - \psi_p)^2/(2\psi_{\sigma1}^2)] + I_{h2}\exp[-(\psi - \psi_p)^2/(2\psi_{\sigma2}^2)]$, yielding $I_0 = (1.76 \pm 0.18) \times 10^{11}$, $I_{h1} = (8.2 \pm 1.6) \times 10^{11}$, $\psi_p = 50.7° \pm 0.1°$, $\psi_{\sigma1} = 1.03° \pm 0.21°$, $I_{h2} = (5.5 \pm 1.6) \times 10^{11}$ and $\psi_{\sigma2} = 3.18° \pm 0.53°$, with $R^2 = 0.996$. The double Gaussian function is necessary to fit the more pronounced Bragg-like peak before release. In Fig. 2j, after release this sharper peak has disappeared, so we fit the azimuthal line shape at intermediate q using only a single Gaussian function, as before, yielding $I_0 = (0.46 \pm 0.19) \times 10^{11}$, $I_h = (1.72 \pm 0.18) \times 10^{11}$, $\psi_p = 51.2° \pm 0.4°$ and $\psi_\sigma = 6.8° \pm 1.0°$, with $R^2 = 0.96$. Thus, the initial sharper Bragg-like nature of this peak before release has been markedly reduced long after release as a consequence of the Brownian fluctuations of the tiles in the P2 system.

Tracking positions and orientations of darts in PSDMs. We wrote a customized particle-tracking routine in Mathematica (Wolfram Research, version 11.2) to track the centroids and pointing directions of the five darts in a fluctuating PSDM and reveal the local symmetry-breaking configurational fluctuations of this motif caused by Brownian excitations (Supplementary Video 4, Extended Data Fig. 5). This routine thresholds, binarizes and fills the five dart tiles. A combination of the *MorphologicalComponents*, *EdgeDetect* and *ImageCorners* functions are used to calculate the centroids of each of the five darts and to locate their three convex vertices (Extended Data Fig. 5a). The centroids are sorted anticlockwise and connected as motif vertices to form a convex pentagon (blue lines), the area and internal angles of which are computed in each frame. The centroids and convex vertices are used to calculate the pointing directions of the darts (red arrows displayed at motif vertices). To correct for a slight long-time drift in the trajectories of the darts, we determine an average position of the centroids of all five darts in each frame, and the individual trajectories of the darts are then calculated relative to this collective frame (Extended Data Fig. 5b). These trajectories exhibit an anisotropy that reflects the underlying five-fold nature of the PSDMs. Because a PSDM has more highly corrugated edges, it is more strongly rotationally coupled to the surrounding tiles than is a PSKM, so it does not exhibit collective rotational hopping motion (as was observed for PSKMs).

Calculations and fits of the distributions of areas and internal angles of fluctuating PSDMs. We calculated normalized probability distributions on the basis of the recorded set of areas A_{PSDM} and internal angles β_{PSDM} of the pentagons of connected motif vertices given by the centroids of the five darts using all frames in Supplementary Video 4 (Extended Data Fig. 5c, d). Because these five darts cannot be compressed below a limiting area A_0, which corresponds to them touching, the area distribution $p_{A,PSDM}$ is asymmetric and exhibits a strict cut-off at low areas; we fit it using a three-parameter normalized log-normal distribution (to simplify notation, we take $A = A_{PSDM}$ and $p_A = p_{A,PSDM}$):

$$p_A = \begin{cases} \dfrac{exp(-\{ln[(A-A_0)/\delta A]\}^2/2\gamma^2)}{\sqrt{2\pi}\,\gamma(A-A_0)} & \text{for } A > A_0 \\ 0 & \text{for } A \le A_0 \end{cases}$$

The results of the fit are displayed in Extended Data Fig. 5c as the black line, and the values of the fit parameters obtained are $A_0 = 52 \pm 7\,\mu m^2$, $\delta A = 22 \pm 7\,\mu m^2$ and $\gamma = 0.17 \pm 0.05$ ($R^2 = 0.994$). Thus, the PSDM exhibits substantial area fluctuations caused by Brownian excitations, reflecting local equilibrium density fluctuations of this particular motif within the P2 quasi-crystal system. These Brownian excitations also lead to local symmetry-breaking fluctuations of the PSDM, as revealed by the sequence of differently shaped blue pentagons in Supplementary Video 4 and Extended Data Fig. 5a; over time, this fluctuating blue pentagon exhibits distortions away from a perfect regular pentagon. To quantify these distortions, we calculate the normalized probability distribution of internal angles $p_{\beta,PSDM}$ of this fluctuating blue pentagon (Extended Data Fig. 5d). We fit this distribution using a Gaussian function, yielding a mean of 69.3° and a standard deviation of 17.5° ($R^2 = 0.99$). A direct calculation using the un-binned list of internal angles gives similar results (mean, 72.0°; standard deviation, 15.9°), irrespective of the binning parameters used to create the distribution. Thus, the time-averaged configuration of darts within a PSDM does reflect the basic five-fold symmetry of this motif through the mean internal angle of 72°, corresponding to a regular pentagon. Nevertheless, Brownian excitations cause large local symmetry-breaking fluctuations in the instantaneous configurations of dart tiles about this symmetric mean, leading to a relatively large and readily observable standard deviation in the distribution of these internal angles.

Melting kite and dart motifs in a P2 quasi-crystal. In Fig. 4, we fit the measured $\phi_A(d)$ at certain fixed times after release t using a modified Fermi-like function to capture the evolution of diffusive dynamics associated with melting as the P2 quasi-crystal expands to occupy empty space after a confining wall is removed: $\phi_A(d) = \phi_A^*/\{1 + \exp[(d - d_0)/L]\}$, where ϕ_A^* is the plateau area fraction in the dense region (to the left), d_0 is a reference distance that roughly defines the position of the melting front and L is a characteristic length scale associated with the width of the melting front. We use similar Fermi-like fitting forms to fit data for the d-dependent area fractions of PSKMs ($\phi_{A,KM}$) and PSDMs ($\phi_{A,DM}$). All fitting parameters for ϕ_A (Fig. 4d), $\phi_{A,KM}$ (Fig. 4e) and $\phi_{A,DM}$ (Extended Data Fig. 6), at the certain t are shown in Extended Data Table 1.

Determining the melting area fraction of the P2 tile system $\phi_{A,melt}$. In Fig. 4i, we fit all data for $\phi_{A,KM}(\phi_A) > 0$ using linear functions. The slopes of $\phi_{A,KM}(\phi_A)$ 24 h, 48 h and 60 h after release are 1.30 ± 0.04, 2.26 ± 0.14 and 2.42 ± 0.53,

respectively, and the intercepts with the ϕ_A axis are 0.544 ± 0.006, 0.633 ± 0.004 and 0.647 ± 0.009, respectively. At $t = 48$ h and $t = 60$ h after release, these intercepts are approximately the same, so we estimate that the area fraction at the melting point of the P2 quasi-crystal is $\phi_{A,melt} \approx 0.65 \pm 0.02$. This value is associated with the melting of PSKMs; larger-scale superstructures of motifs become ill-defined when the fundamental motif structures become disorganized.

Immediate melting of P2 kite and dart tiles at a lower area fraction. We also pre-assembled kite and dart tiles at a uniformly lower area fraction of $\phi_A \approx 0.53$ by creating even larger spaces between adjacent tiles in a Penrose P2 pattern mask design, but still enclosing all tiles using large confining walls. As the kite and dart tiles are released after adding the RSD, the P2 pattern melts and becomes a disordered liquid-like phase because the tiles have more empty space in which they can rotate and translate, even as they stay in the monolayer. This immediate melting behaviour at uniform $\phi_A \approx 0.53$ is consistent with the higher melting area fraction $\phi_{A,melt} \approx 0.65 \pm 0.02$ that we determined above for the unconfined P2 system. Although specific tile area fractions were designed into the mask patterns, the area fraction of tiles printed using a particular pattern design can be reduced by lowering the exposure dose in the stepper down to as low as about 150 mJ cm^{-2}. At even lower doses below this value, the cross-linking within the tiles is compromised and vertex rounding becomes extreme. This limited control over the area fraction of tiles at the printing stage could potentially be useful because it could eliminate the need to design a new mask pattern.

Additional discussion. Although different top-down lithographic methods have been used to produce a wide variety of custom shape-designed colloidal particles[9–11], these processes typically yield stable bulk dispersions of desired shapes in a liquid. These shapes have subsequently been used in depletion-driven[31–33] and capillary-driven[34,35] self-assembly experiments, providing insights into how features in the shapes of particles can influence self-assembly. For instance, lithographic mutations of sub-particle features of colloidal chiral C-shapes that resemble proteins, known as proteoids, have been used to control the entropic hierarchical self-assembly of dimer crystals under slow crowding in two dimensions as roughness-controlled depletion attractions keep the fluctuating monolayer intact[36]. Nevertheless, the sizes of crystallites of such self-assembled structures are typically small, and a high defect density often accompanies this type of self-assembly. Although such earlier studies have provided important insights into the role of core shape and entropy in self-assembly of gradually crowded systems, including protein crystallization, these limitations have precluded the widespread use of such self-assembled structures. By contrast, litho-PAMs can be used to produce and study complex fluctuating multiscale systems of mobile tiles at high tile densities, as we have demonstrated using optical stepper lithography and optical microscopy. Thus, litho-PAMs are very different from previous 2D self-assembly experiments in which monolayers of shape-designed lithographic colloids have been randomly deposited at dilute area fractions in a monolayer and subsequently concentrated slowly using gravitational sedimentation in the presence of roughness-controlled depletion attractions[25]. The tilted-cell approach creates a gradient in particle area fraction throughout; by contrast, the litho-PAM method provides a uniform area fraction over a very large surface area. Thus, pre-assembly at uniform ϕ_A avoids potential issues of out-of-equilibrium jamming during crowding as well as the spatial gradients in ϕ_A that are inherent in the tilted-cell method of 2D self-assembly.

Achieving a monolayer of fluctuating kite and dart tiles in a Penrose P2 quasi-crystalline pattern[37] that is defect-free over large areas and in which single-particle and multi-particle collective dynamics can be readily visualized advances the experimental science of multiscale complex systems of mobile colloidal particles. Moreover, this achievement demonstrates that new equilibrium phases composed of many differently shaped and configured building blocks can be readily produced using top-down parallel fabrication methods, circumventing bottom-up self-assembly methods and serial directed-assembly methods. Litho-PAMs can even enable the creation and study of different individual ground-state configurations for systems that have degenerate ground states with the same free energy yet different polymorphic organizations of tiles. Such degeneracy would usually preclude simple self-assembly methods from producing different desired ground-state configurations, because different ground-state polymorphs with essentially the same energy would nucleate and grow in different local regions[38].

Until now, attempts to create a large-scale fluctuating Penrose quasi-crystal of mobile colloidal tiles, which are organized with five-fold symmetry and have effectively hard interactions, have not succeeded. Further, no experimental assembly method has been able to produce a large-scale fluctuating colloidal version of the five-fold Penrose P2 quasi-crystal[39] that can be observed in real-space with particle-scale resolution until now. In a previous report of a self-assembled soft-matter quasi-crystal of nanoparticles that were crowded through evaporation, the crystallites were limited to 12-fold[7,8] and 18-fold symmetry[8]. It has been proposed and investigated in simulations[40–42] that a Brownian Penrose quasi-crystal could be self-assembled by introducing selective site-specific edge–edge interactions between the tiles; however, this has not been realized experimentally, probably because of the substantial complexity of creating the required variety of selective edge–edge interactions on real particles. By contrast, the litho-PAM approach overcomes these limitations, enabling us to produce the quasi-crystal presented here. Moreover, we have visualized the equilibrium fluctuations of this quasi-crystal, including heterogeneous collective dynamics of certain motifs, as well as melting by slowly lowering the tile density. Our ability to observe mobile tiles fluctuating in equilibrium using optical microscopy has enabled us to identify the hallmarks of the pentatic liquid quasi-crystalline phase of matter, which is analogous to the fluctuating hexatic phase of mobile hard disks subject to Brownian excitations. This provides in situ dynamic information that cannot be obtained by studying a static solid film of nanoparticles using electron beam microscopy after evaporating the continuous liquid phase.

Fundamental questions about spatio-temporal dynamics arise when considering complex multiscale materials composed of mobile building blocks, such as the P2 quasi-crystal system. Traditional models of phase behaviour and equilibrium fluctuations, such as Onsager-like cage models, do not necessarily apply to multiscale systems that have a large variety of dynamic motifs and patterns that can be hundreds or more times larger than the smallest particles. Understanding equilibrium and non-equilibrium dynamics of multiscale systems that have differently organized structures at different length scales, including those well beyond the colloidal scale, remains a challenge theoretically and experimentally. For example, directly applying the notions of Kosterlitz–Thouless theory[43] to a fluctuating P2 quasi-crystal is challenging; yet, on the basis of our observations, doing so would probably show the emergence of liquid-crystalline-like features in correlation functions and order parameters at smaller length scales for values of ϕ_A between the disordered liquid state at low ϕ_A and the full-tiling limit as $\phi_A \to 1$. Just as disks can form a hexatic liquid-crystal phase[27] over a certain range of ϕ_A, our observations indicate that liquid-crystal-like modulations in the Fourier transforms develop for Brownian quasi-crystals of hard P2 tiles at intermediate wavenumbers q. Yet, we also find that the degree of fluctuation-induced smearing of quasi-crystal Bragg peaks at our smallest observable q, corresponding to the largest length scales, is not as large as the degree of fluctuation-induced smearing at intermediate q. These features differentiate the Fourier transforms of the fluctuating pentatic liquid quasi-crystalline phase from the q-independent perfect Bragg peaks associated with crystallography of ideal static quasi-crystals constructed geometrically[44]. The $g_{MSO\text{-}PCF}(r)$ that we developed (see Supplementary Methods) is a first attempt at quantifying superstructural orientational order in hierarchically organized multiscale materials in real space, and this concept can be further broadened and generalized. This brings up an important theoretical question for multiscale materials, which can have different symmetries and potentially incommensurate organizational structures at different length scales, of what is meant by short-, intermediate- and long-range order. Using the Brownian P2 quasi-crystal as an example, we believe that it may be necessary to correlate the type and degree of order with a range of length scales (or equivalently range of q) in multiscale materials, and possibly even spatial locations; any new theory would seek to couple these differently structured regions together self-consistently. Furthermore, a theoretical exploration of melting in hierarchically organized systems of differently shaped fluctuating building blocks could enable a direct comparison with our experimental results of melting. At a first level, such a theoretical exploration could be based around a multi-body free energy that is entirely entropic in origin, by considering the entropy of allowed microstates of positions and orientations of all tiles subject to the constraints of non-overlap.

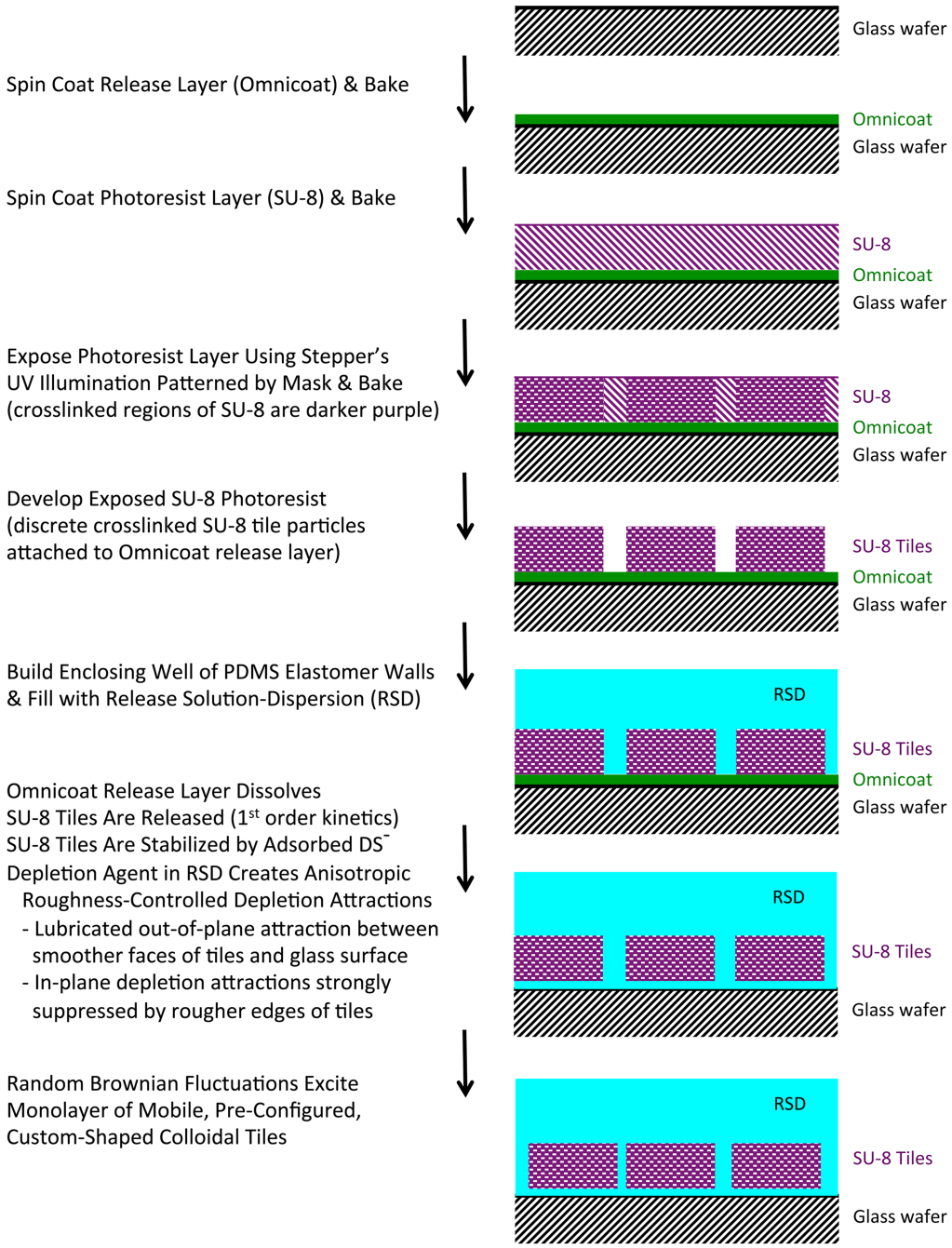
For classic solid-state atomic quasi-crystals, the notion of phason strain[45] can be used to explain real-space experimental observations that positions of certain atoms do not conform to a perfect ideal quasi-crystal structure[46]. To analyse these images, quasi-crystal tiles are decorated with dots and lines to represent atoms and the bonds between atoms, respectively. Alterative tilings, which are not perfect Penrose quasi-crystals and do not follow standard matching rules everywhere but still fill space, can be made with these decorated tiles, thereby creating different local isomorphs. These alternative tilings typically lead to the phenomenon of phason strain when atom and bond decorations in adjacent tiles do not match up. However, in the case of our quasi-crystal, the hard tiles are the fundamental particles, so there are neither atoms nor covalent bonds between constituent tiles. Thus, atomic interpretations of phason strain, which have been used to explain phenomena in solid-state quasi-crystals with very strong attractive interactions between constituent atoms, are not directly applicable to systems of dispersed hard tiles. Instead, a description assuming an entirely entropic free energy, as previously stated, would be most appropriate for predicting the equilibrium properties and fluctuations of such systems. It would be possible to print alternative local isomorphs that do not necessarily follow matching rules for perfect Penrose quasi-crystals using the litho-PAM method, but the free energies of tile systems containing these isomorphs may be degenerate because there are no substantial in-plane attractive interactions between the tiles. In the future, litho-PAMs could be used to introduce different types of defect in the tilings, such as dislocations and

disclinations, which could have a substantial effect on a free energy of the tiles that is primarily entropic in origin.

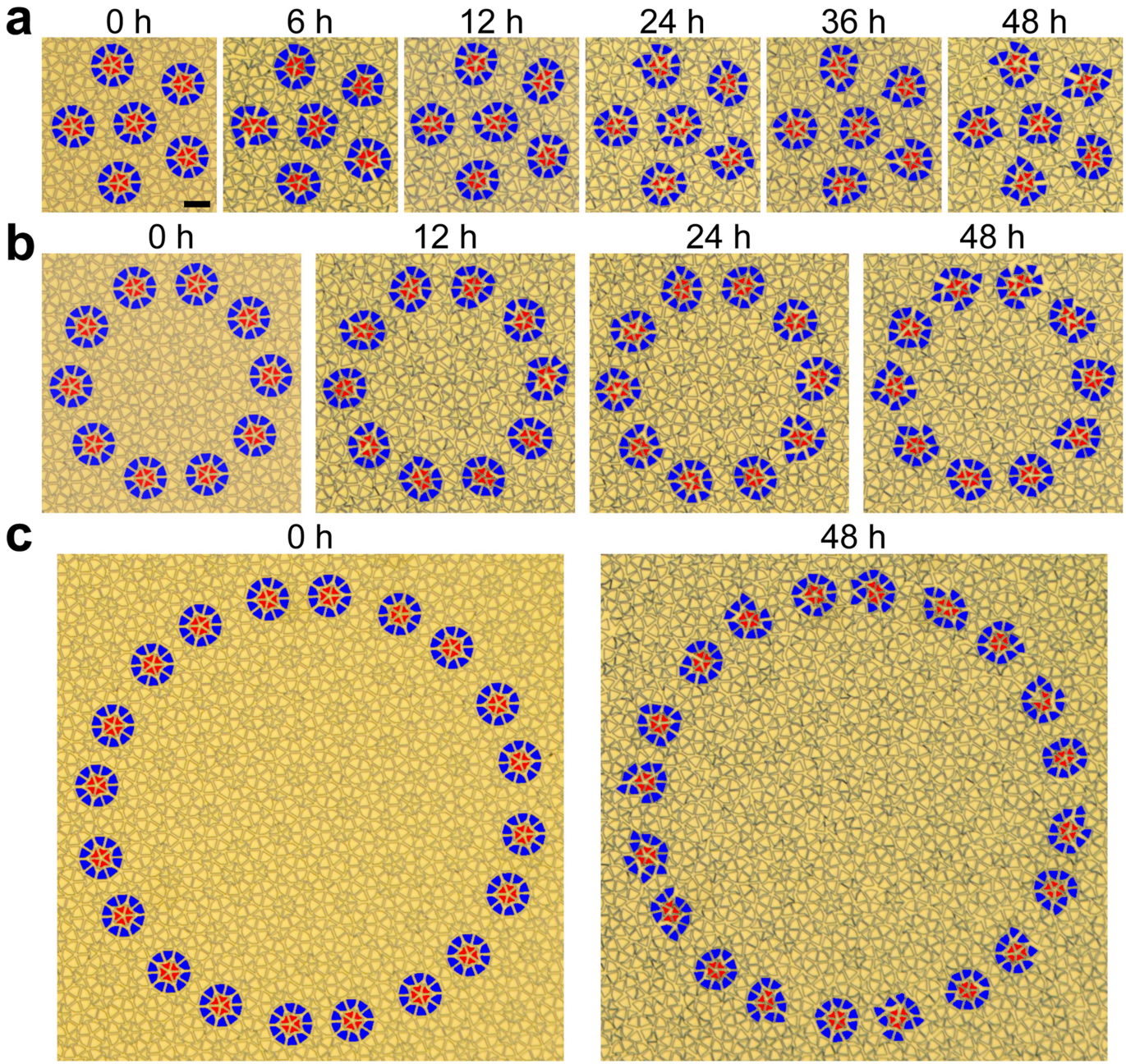
It is interesting to consider the motifs and superstructures of motifs that we have identified through configurations of P2 tiles at different scales in light of previous work on coverings and quasi-unit cells. Overlapping decorated decagonal tiles has been shown[47] to be a useful method of generating pentagonal quasi-crystalline tilings according to prescribed geometrical rules that govern the overlap. These overlapping decagonal systems were referred to as 'coverings', to distinguish them from 'tilings', which typically do not have overlap. Maximizing the density of decorated quasi-unit cells was introduced[41] as a simple yet powerful, general method of generating quasi-crystalline tilings. Using this quasi-unit cell approach, it is possible to explain, for example, the formation of $Al_{72}Ni_{20}Co_8$ quasi-crystals observed[48] using high-angle annular dark-field imaging. The C-clusters (which represent an energetically preferred low-energy atomic cluster, implying strong attractive interactions between atoms compared to k_BT) of the quasi-unit cells[41] are overlapped, and when their density is maximized a Penrose tiling is produced. Thus, in these previous works, overlap of decagons or of quasi-unit cells is needed to produce quasi-crystalline coverings or tilings. By contrast, our system, which is composed of non-overlapping kite and dart tiles in a monolayer at an area fraction of less than unity, is formally neither a covering nor a close-packed tiling (which would correspond to $\phi_A = 1$).

Data availability. The data shown in the figures and that support the findings of this study are available from the corresponding author on reasonable request.

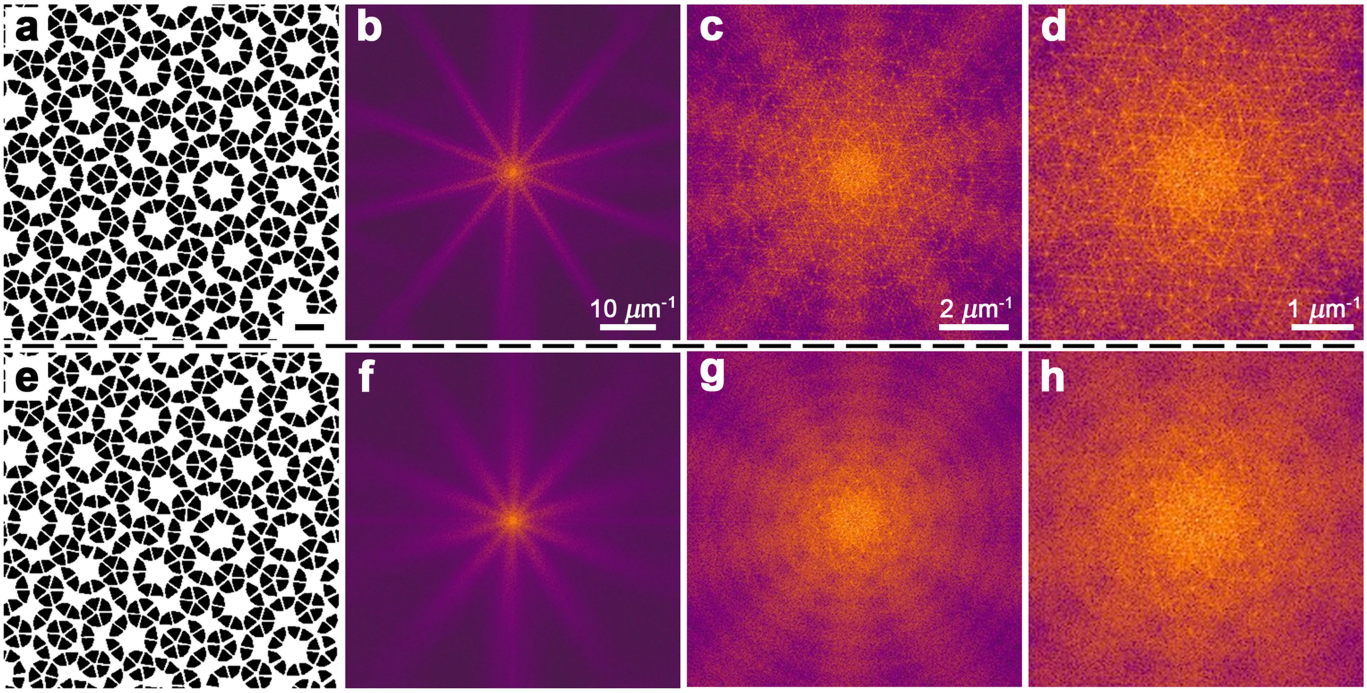
31. Wang, P.-Y. & Mason, T. G. Colloidal lock-and-key dimerization reactions of hard annular sector particles controlled by osmotic pressure. *J. Am. Chem. Soc.* **137**, 15308–15314 (2015).
32. Zhao, K., Bruinsma, R. & Mason, T. G. Entropic crystal–crystal transitions of Brownian squares. *Proc. Natl Acad. Sci. USA* **108**, 2684–2687 (2011).
33. Zhao, K. & Mason, T. G. Twinning of rhombic colloidal crystals. *J. Am. Chem. Soc.* **134**, 18125–18131 (2012).
34. Lewandowski, E. P., Bernate, J. A., Tseng, A., Searson, P. C. & Stebe, K. J. Oriented assembly of anisotropic particles by capillary interactions. *Soft Matter* **5**, 886–890 (2009).
35. Cavallaro, M., Botto, L., Lewandowski, E. P., Wang, M. & Stebe, K. J. Curvature-driven capillary migration and assembly of rod-like particles. *Proc. Natl Acad. Sci. USA* **108**, 20923–20928 (2011).
36. Wang, P.-Y. & Mason, T. G. Dimer crystallization of chiral proteoids. *Phys. Chem. Chem. Phys.* **19**, 7167–7175 (2017).
37. Gardner, M. *Penrose Tiles to Trapdoor Ciphers* (W. H. Freeman, New York, 1989).
38. Zhao, K. & Mason, T. G. Shape-designed frustration by local polymorphism in a near-equilibrium colloidal glass. *Proc. Natl Acad. Sci. USA* **112**, 12063–12068 (2015).
39. Caspar, D. L. D. & Fontano, E. Five-fold symmetry in crystalline quasicrystal lattices. *Proc. Natl Acad. Sci. USA* **93**, 14271–14278 (1996).
40. Dotera, T., Oshiro, T. & Ziherl, P. Mosaic two-lengthscale quasicrystals. *Nature* **506**, 208–211 (2014).
41. Steinhardt, P. J. & Jeong, H.-C. A simpler approach to Penrose tiling with implications for quasicrystal formation. *Nature* **382**, 431–433 (1996).
42. Dontabhaktuni, J., Ravnik, M. & Zumer, S. Quasicrystalline tilings with nematic colloidal platelets. *Proc. Natl Acad. Sci. USA* **111**, 2464–2469 (2014).
43. Kosterlitz, J. M. & Thouless, D. J. Ordering, metastability and phase transitions in two-dimensional systems. *J. Phys. Chem.* **6**, 1181–1203 (1973).
44. Steurer, W. & Deloudi, S. *Crystallography of Quasi-Crystals* (Springer, Berlin, 2009).
45. Socolar, J. Phason strain in quasicrystals. *J. Phys. Colloques* **47**, 217–226 (1986).
46. Li, F. H., Pan, G. Z., Huang, D. X., Hashimoto, H. & Yokota, Y. Phason-strain identification for quasicrystals by high-resolution electron microscopy. *Ultramicroscopy* **45**, 299–305 (1992).
47. Gummelt, P. Penrose tilings as coverings of congruent decagons. *Geom. Dedicata* **62**, 1–17 (1996).
48. Saitoh, K., Tsuda, K., Tanaka, M., Kaneko, K. & Tsai, A. P. Structural study of an $Al_{72}Ni_{20}Co_8$ decagonal quasicrystal using the high-angle annular dark-field method. *Jpn. J. Appl. Phys.* **36**, L1400–L1402 (1997).



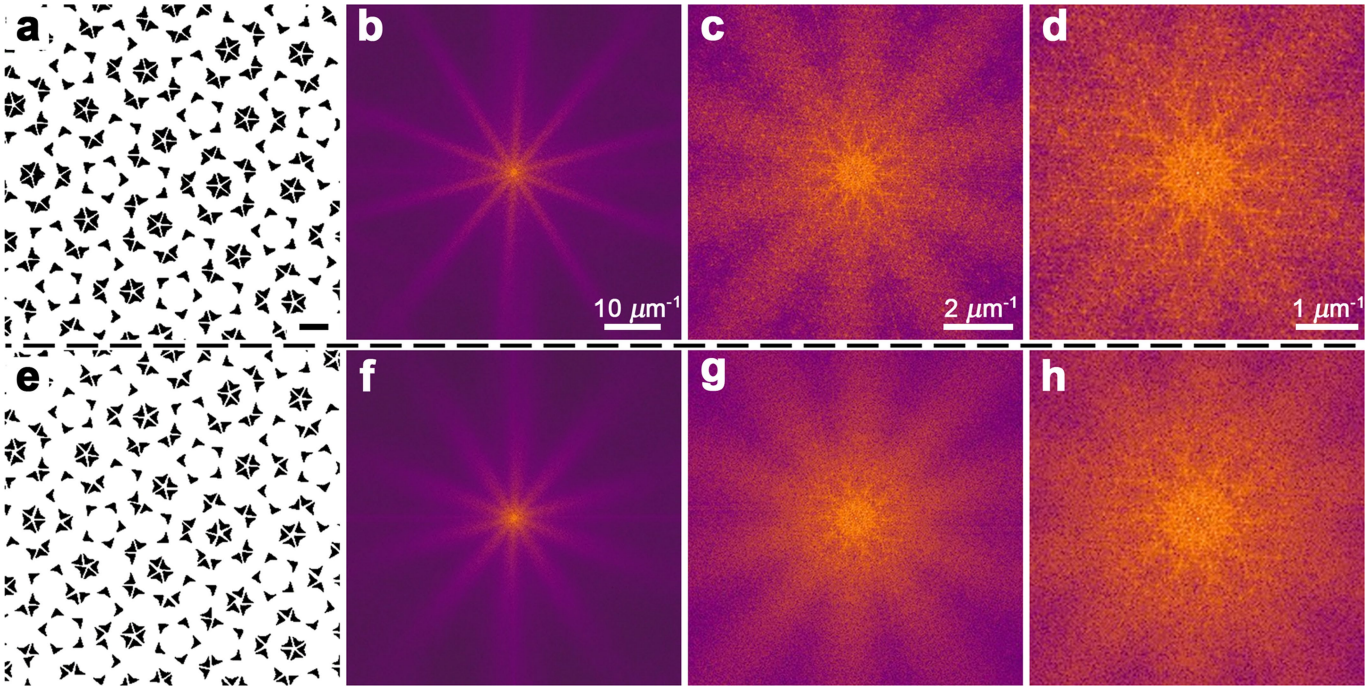
Extended Data Fig. 1 | Detailed sequence of steps for fabricating and releasing a dense litho-PAM of mobile shape-designed tiles. The tiles (purple), which are composed of cross-linked polymeric SU-8 photoresist, are released after exposure and development from the glass wafer (black hatching) using an RSD (blue) that contains a release agent (TMAH) to dissolve the Omnicoat release layer (green), a stabilizing agent (SDS) that adsorbs onto released tiles and prevents their aggregation, and a depletion agent (nanoscale polystyrene spheres) that strongly inhibits the released tiles from leaving the monolayer as a consequence of Brownian excitations (steps are shown as side views).



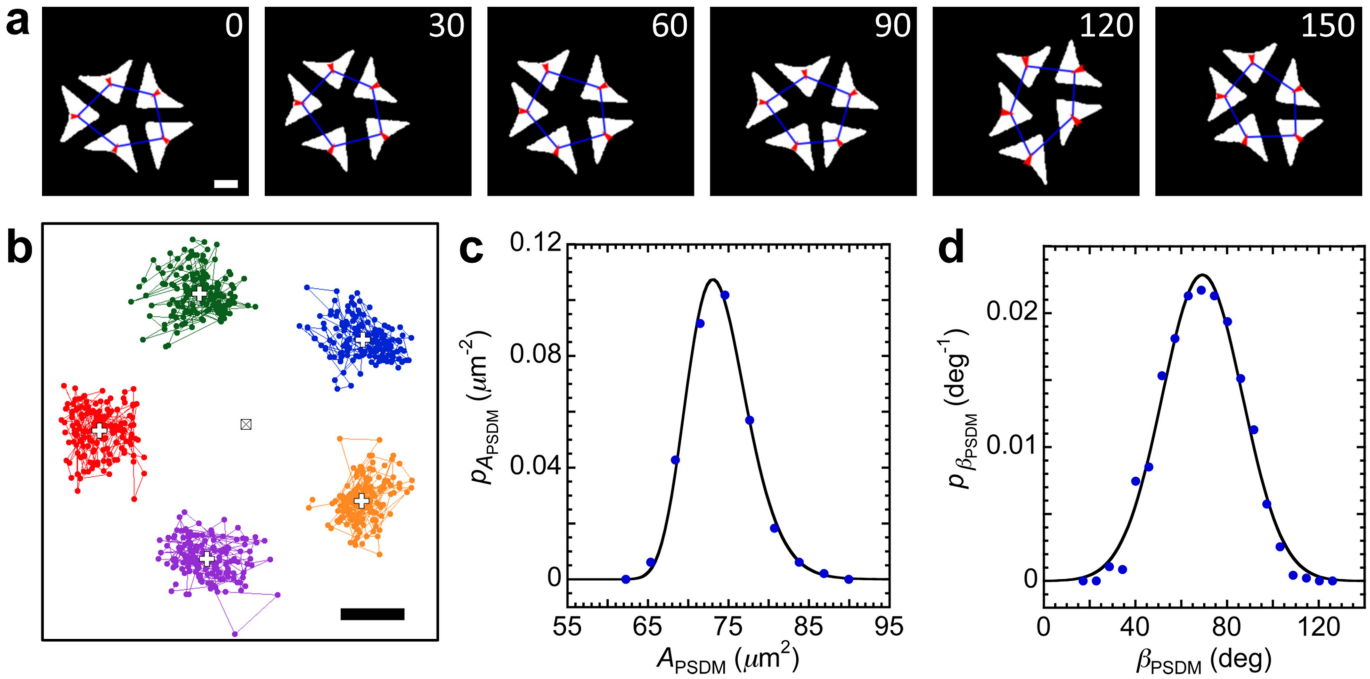
Extended Data Fig. 2 | Examples of organized superstructural sets of motifs extending over different length scales in the P2 quasi-crystal before and after release. a, Regular centre-filled pentagonal superstructural set of one central and five outer wheel motifs (each wheel motif consists of ten kites (filled blue) and five darts (filled red)) before release (0 h, leftmost) and after release (6 h, 12 h, 24 h, 36 h and 48 h, left to right). **b**, Regular decagonal superstructural set of ten wheel motifs before release (0 h, leftmost) and after release (12 h, 24 h and 48 h, left to right). **c**, Regular icosagonal (that is, a 20-sided polygon) superstructural set of 20 wheel motifs before release (0 h, left) and after release (48 h, right). Scale bar (20 μm) shown in **a** is the same for all images.



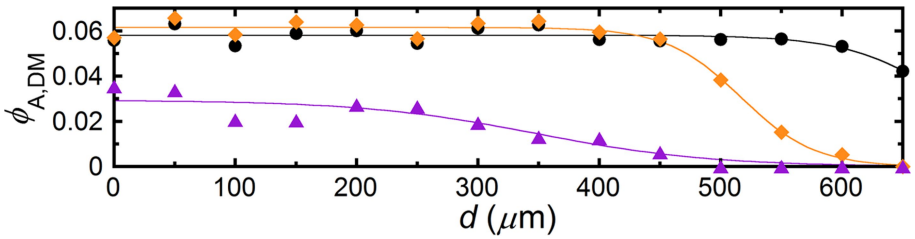
Extended Data Fig. 3 | Restructuring of Penrose kite tiles in a P2 quasi-crystal after release. **a**–**d**, Before release; **e**–**h**, after release. **a**, Kite tiles are separated by post-processing of the micrograph image from Fig. 2a. Scale bar, 20 μm. **b**, Effective scattering pattern, given by the Fourier transform intensity of **a**, showing ten rays extending from the centre to high scattering wavenumbers q. **c**, Central region of **b**, magnified by a factor of about six, revealing Bragg peaks at intermediate and low q, corresponding to large distances. **d**, Central region of **c**, magnified by a factor of about two, revealing Bragg peaks at very low q associated with superstructural ordering of motifs of kite tiles over large length scales. **e**, Brownian fluctuations near equilibrium (48 h after release) have caused kite tiles to deviate from the original perfect quasi-crystal order seen in the unreleased structure; kites are separated from Fig. 2e. **f**, Fourier transform intensity of **e**. Rays have broadened azimuthally as a consequence of Brownian fluctuations. **g**, **h**, Close-ups of **f** over the same q ranges as for **c** and **d**, respectively, showing the smearing of Bragg peaks into ten-fold azimuthal intensity modulations, reminiscent of liquid-crystalline materials, indicating a retention of quasi-crystalline orientational order. Intensity colour scale is the same as in Fig. 2k.



Extended Data Fig. 4 | Restructuring of Penrose dart tiles in a P2 quasi-crystal after release. **a**–**d**, Before release; **e**–**h**, after release. **a**, Dart tiles are separated by post-processing of the micrograph image from Fig. 2a. Scale bar, 20 μm. **b**, Effective scattering pattern, given by the Fourier transform intensity of **a**, showing ten rays extending from the centre to high scattering wavenumbers q. **c**, Central region of **b**, magnified by a factor of about six, revealing Bragg peaks at intermediate and low q, corresponding to large distances. **d**, Central region of **c**, magnified by a factor of about two, revealing Bragg peaks at very low q associated with superstructural ordering of motifs of dart tiles over large length scales. **e**, Brownian fluctuations near equilibrium (48 h after release) have caused dart tiles to deviate from the original perfect quasi-crystal order seen in the unreleased structure; darts are separated from Fig. 2e. **f**, Fourier transform intensity of **e**. Rays have broadened azimuthally as a consequence of Brownian fluctuations. **g**, **h**, Close-ups of **f** over the same q ranges as for **c** and **d**, respectively, showing the smearing of Bragg peaks into ten-fold azimuthal intensity modulations, reminiscent of liquid-crystalline materials, indicating a retention of quasi-crystalline orientational order. Intensity colour scale is the same as in Fig. 2k.



Extended Data Fig. 5 | Tracking anisotropic, bounded Brownian fluctuations of darts in a PSDM. a, Filled and thresholded optical micrographs of darts in individual video frames (frame number in upper right) are overlayed with blue pentagons with vertices at the centroids of the darts (see Methods, Supplementary Video 4). Red arrows at the vertices of the blue pentagon denote the pointing directions of the darts. The shapes of the blue pentagons fluctuate over time and at any given instant can deviate substantially from a regular pentagon as a consequence of Brownian excitations of the P2 system. Actual time between frames is 720 s. Scale bar, 3 μm. **b**, Trajectories of the centroids of five darts in a PSDM over a duration of 32 h, after correcting for a slight long-time drift of the entire motif. The time-average position of each dart is denoted by a plus symbol overlaid on each trajectory; the centre of the PSDM is given by crossed box symbol at the centre. These trajectories have non-circular shapes, indicating that the bounded Brownian motion of the darts is anisotropic, reflecting the local five-fold time-averaged quasi-crystal symmetry of the motif. Considering an ensemble average over all five darts, standard deviations of step size distributions projected along directions between the centre of the motif and the time-averaged centroids of the five darts are 1.3 ± 0.2 times less than standard deviations of step size distributions projected perpendicular to these five directions. This small yet detectable anisotropy in the bounded diffusion of the darts reflects the underlying time-averaged five-fold symmetry of their local quasi-crystal environment. Scale bar, 3 μm. **c**, **d**, Normalized probability distributions of the calculated area A_{PSDM} (**c**) and internal angles β_{PSDM} (**d**) of the fluctuating blue pentagons in Supplementary Video 4 and **a**. Black lines, fits using a log-normal distribution (**c**) and a Gaussian distribution (**d**); see Methods for functional forms and fit parameters.



Extended Data Fig. 6 | Melting of PSDMs in an unconfined P2 quasi-crystal. The area fraction of PSDMs $\phi_{A,DM}$ (red darts in Fig. 4a–c) decays to zero at larger distances d, towards the direction where the quasi-crystal is not confined (right). The motif melting front moves from right to left over time: black circles, 4 h after release; orange diamonds, 24 h after release; purple triangles, 48 h after release. Fits are using a Fermi-like function (see Methods); fit parameters are given in Extended Data Table 1.

Extended Data Table 1 | Fit parameters of the Fermi-like functions that describe the dependence of ϕ_A, $\phi_{A,KM}$ and $\phi_{A,DM}$ on d during melting of the P2 quasi-crystal

Data Fit	Fitting Parameter	t = 4 h	t = 24 h	t = 48 h
$\phi_A(d)$	ϕ_A*	0.770 ± 0.001	0.768 ± 0.003	0.762 ± 0.002
	d_0 (μm)	846 ± 16	744 ± 9	773 ± 5
	L (μm)	98 ± 7	95 ± 7	162 ± 5
	R^2	0.994	0.994	0.999
$\phi_{A,KM}(d)$	$\phi_{A,KM}$*	0.314 ± 0.003	0.285 ± 0.005	0.269 ± 0.013
	$d_{0,KM}$ (μm)	750 ± 35	567 ± 6	387 ± 16
	L_{KM} (μm)	79 ± 22	49 ± 6	66 ± 13
	R^2	0.921	0.989	0.981
$\phi_{A,DM}(d)$	$\phi_{A,DM}$*	0.058 ± 0.001	0.062 ± 0.001	0.029 ± 0.003
	$d_{0,DM}$ (μm)	691 ± 7	517 ± 5	348 ± 29
	L_{DM} (μm)	41 ± 8	30 ± 4	72 ± 24
	R^2	0.953	0.995	0.953

Fits are shown in Fig. 4d, e and Extended Data Fig. 6; see Methods for the fitting functions and definitions of the fit parameters. R^2, correlation coefficient; t, time after initiation of release.

LETTER

https://doi.org/10.1038/s41586-018-0463-x

Climate–induced changes in continental–scale soil macroporosity may intensify water cycle

Daniel R. Hirmas[1]*, Daniel Giménez[2], Attila Nemes[3], Ruth Kerry[4], Nathaniel A. Brunsell[5] & Cassandra J. Wilson[5,6]

Soil macroporosity affects field-scale water-cycle processes, such as infiltration, nutrient transport and runoff[1,2], that are important for the development of successful global strategies that address challenges of food security, water scarcity, human health and loss of biodiversity[3]. Macropores—large pores that freely drain water under the influence of gravity—often represent less than 1 per cent of the soil volume, but can contribute more than 70 per cent of the total soil water infiltration[4], which greatly magnifies their influence on the regional and global water cycle. Although climate influences the development of macropores through soil-forming processes, the extent and rate of such development and its effect on the water cycle are currently unknown. Here we show that drier climates induce the formation of greater soil macroporosity than do more humid ones, and that such climate-induced changes occur over shorter timescales than have previously been considered—probably years to decades. Furthermore, we find that changes in the effective porosity, a proxy for macroporosity, predicted from mean annual precipitation at the end of the century would result in changes in saturated soil hydraulic conductivity ranging from −55 to 34 per cent for five physiographic regions in the USA. Our results indicate that soil macroporosity may be altered rapidly in response to climate change and that associated continental-scale changes in soil hydraulic properties may set up unexplored feedbacks between climate and the land surface and thus intensify the water cycle.

At a continental scale, the biogeochemical and mechanical processes responsible for the formation of macropores[2] (for example, soil biological activity, tillage practices and the natural formation of soil aggregates) are mediated by climatological factors, such as precipitation and temperature, that largely control soil moisture and energy fluxes. By inference, therefore, the formation of macropores in the soil must also be mediated by climate. Given the control of soil macroporosity over water cycling between the land and atmosphere, predicting its response to future climate scenarios is crucial. This prediction is currently impossible, however, because the degree and rate at which climate influences soil macroporosity are unknown.

Moreover, research examining the response of the land surface to climate change has commonly assumed that hydraulic properties of the soil that control water flux through the vadose zone, such as saturated hydraulic conductivity, remain static over timescales relevant to climate simulations[5,6]. Climosequence studies, by contrast, indicate that soils aggregate into distinct structural units under the prevailing soil climatic conditions, forming interaggregate macropores through chemical and physical processes—such as the formation of stable associations of organic matter, clay and silt particles[7] and wet–dry[8] and freeze–thaw cycles[9]—that can affect macroporosity seasonally. Biopores created via root penetration[10] and faunal burrowing[11] are also influenced by climate. The overall rate at which macropores develop from the combination of these processes is unknown, although it could be as rapid as 15 years in organic soils[12].

Here we investigate the implicit assumption that soil macroporosity—and, thus, soil hydraulic properties—are static over the timescales of end-of-century model simulations by examining continental-scale effects of climate on the development of effective porosity using a large database of soils sampled across the USA over approximately the past 50 years. Although multiple definitions have been proposed for effective porosity, we consider it to be the difference between total porosity and field capacity[13]. Field capacity is the water content of an initially saturated soil that has ceased to drain owing to a balance between soil matric potential and gravitational forces. Because this balance is often assumed to be represented by a soil water pressure potential of −33 kPa, corresponding to pores of approximately 9 μm in diameter, field capacity represents the volume fraction of soil pores smaller than this diameter. As the difference between total porosity and field capacity, therefore, effective porosity is a simple but useful proxy of macroporosity, representing the volume fraction of the largest pores in the soil[13].

Soil data were obtained from the National Cooperative Soil Survey (NCSS) Characterization Database, which contains information on soils sampled and measured by the USA Department of Agriculture's Natural Resources Conservation Service (USDA–NRCS) and cooperating laboratories from the late 1960s to the present. Although other soil databases have been compiled from various sources and research projects, a major advantage of the NCSS database is that soils were analysed by adhering strictly to well documented standard operating procedures, making the soil information highly reliable and thus comparable, despite the widespread sampling dates and geographic distribution of the samples (Extended Data Fig. 1).

Macroporosity is strongly dependent on particle-size distribution and the fraction of soil carbon[14]. To evaluate the influence of climate on effective porosity, therefore, we treated the sand, clay and soil organic matter contents of each selected soil as covariates and calculated the residual effective porosity as the difference between measured effective porosity and the effective porosity predicted from a spatial-error regression model using these covariates. To assess climatic controls of soil macroporosity with depth, we grouped soil samples by morphology into surface layers (A horizons) that occurred within 0–25 cm of the land surface and subsurface layers (B horizons) that occurred within 25–100 cm of the land surface (Extended Data Table 1). To minimize effects from confounding variables, morphological horizons that were designated from field excavations to contain concentrations of aggregating agents, such as carbonate (for example, Bk), or were indurated (for example, Bqm) were not considered in this study. We additionally selected ploughed surface layers (Ap horizons; Extended Data Table 1, Extended Data Fig. 1) to assess effects of climate on the rate of regeneration of macroporosity disturbed by tillage. The adoption of non-inversion tools over the past 50 years led to reduced mechanical disturbance of Ap horizons[15] and facilitated the expression of climate on macroporosity in ecosystems that are uniformly managed and support similar vegetation across climatic zones[16]. An unknown fraction of these Ap samples probably came from irrigated fields; this additional water may have reduced, but not eliminated, soil moisture differences compared to samples unaffected by tillage.

[1]Department of Environmental Sciences, University of California, Riverside, CA, USA. [2]Department of Environmental Sciences, Rutgers University, New Brunswick, NJ, USA. [3]Division of Environment and Natural Resources, Norwegian Institute of Bioeconomy Research, Ås, Norway. [4]Department of Geography, Brigham Young University, Provo, UT, USA. [5]Department of Geography and Atmospheric Science, University of Kansas, Lawrence, KS, USA. [6]Present address: KRNV News 4, Reno, NV, USA. *e-mail: daniel.hirmas@ucr.edu

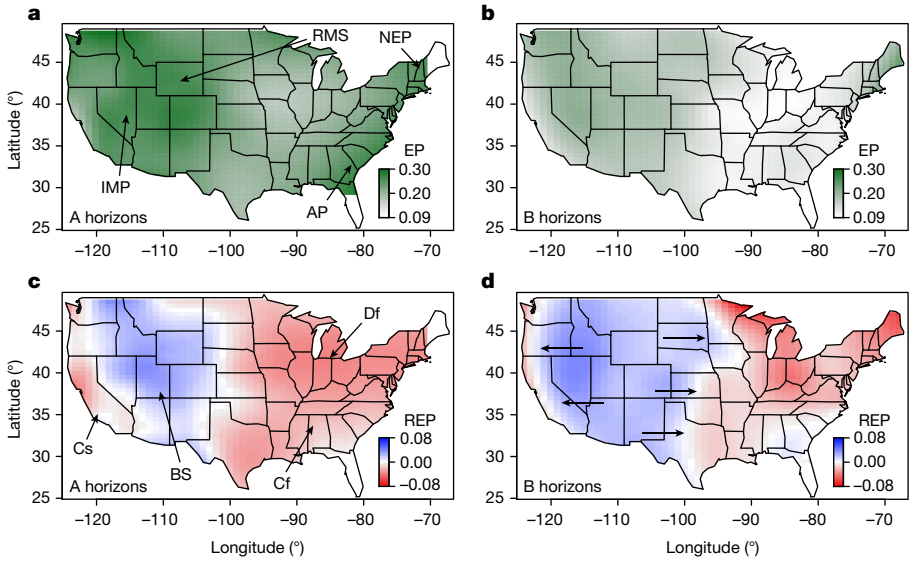


Fig. 1 | Interpolated maps showing continental-scale patterns of macroporosity before and after accounting for non-climatic factors using regression with spatially correlated errors. a–d, Effective porosity (EP; **a**, **b**) and residual effective porosity (REP; **c**, **d**) for the surface (**a**, **c**) and subsurface (**b**, **d**) layers examined in this study. Effective porosity values represent volumetric porosity. Arrows in **a** point to selected physiographic regions: RMS, Rocky Mountain System; NEP, New England province; AP, Atlantic Plain; IMP, Intermontane Plateaus. Arrows in **c** point to selected Köppen–Geiger climatic zones referred to in the text. Arrows in **d** illustrate the lateral extent of residual effective porosity in subsurface layers compared to surface layers (**c**).

Figure 1a, b shows interpolated distributions of effective porosity for surface and subsurface layers. The visual pattern of the effective porosity of a surface layer follows the spatial distribution of physiographic provinces, with high values in the Rocky Mountain System and Intermontane Plateaus in the western half of the conterminous USA and in the New England province and Atlantic Plain in the east. This pattern suggests that the distribution of macroporosity is largely a function of soil texture[13] and possibly mineralogy, both of which are derived from soil parent material. For subsurface layers, the pattern is more muted compared to surface layers, especially in the Atlantic Plain, probably reflecting the illuvial processes that have increased the clay content of the B horizons[17] (Fig. 1b). The enrichment in translocated clay and the lower effective porosity values in subsurface layers also suggest that soil texture is the dominant variable controlling the expression of effective porosity at this scale. However, when the influence of soil texture and organic matter are removed, the distribution follows a more distinct pattern with climate (Fig. 1c, d), with low residual effective porosity of the surface layers occurring in fully humid snow climates (Df in the modified Köppen–Geiger classification[18]), moderate residual effective porosity values in warm temperate climates that either have dry summers (Cs) or are fully humid (Cf), and high residual effective porosity values in arid steppe (BS) climates. These results indicate that drier and warmer climates promote the development of surface-layer macroporosity, whereas more humid and cooler climates restrict the expression of macroporosity. Residual effective porosity in subsurface layers follows the same basic pattern as in surface layers but shows a larger lateral extent of high values in both eastern and western directions (Fig. 1d). This subsurface residual effective porosity pattern resembles the distribution of precipitation timing (see Extended Data Fig. 2) probably owing to the effect of rainfall frequency on the average penetration depth of percolating meteoric water.

We plot the mean residual effective porosity for samples obtained from locations nearest USA Historical Climatology Network (USHCN) stations that recorded observations between 1951 and 2011 against four climatological parameters—mean annual precipitation, mean atmospheric vapour pressure deficit for the months of June, July and August, mean precipitation event magnitude and mean freezing atmospheric temperature frequency—to assess the influence of these variables over soil macroporosity (Fig. 2). Residual effective porosity shows an inverse relationship with both mean annual precipitation and mean precipitation event magnitude and a positive relationship with atmospheric demand for moisture (maximum vapour pressure deficit) in natural surface layers (A horizons). This provides further evidence that the development of macroporosity is restricted under humid climates with larger individual rainfall events than under drier climates. Several factors may be responsible for this trend, including faster turnover rates of macroaggregates in response to more frequent pulses of water[8], which would shift soil pores towards smaller sizes. We observed a positive association with the frequency of freezing temperatures in the atmosphere, suggesting that abiotic mechanisms may also contribute to the development of macroporosity in the surface. Except for freezing frequency, the same relationships between residual effective porosity and climatological variables were observed in subsurface layers (B horizons) but with lower variability, probably because they reflect longer-term averages of atmospheric conditions (owing to their depth) that have led to the development of these horizons.

Because Ap horizons represent similar and often frequent (annual to decadal) soil disturbance across various climate zones, owing to the relative geographic uniformity of agricultural management practices, these surface layers act as a proxy for the rate of macroporosity regeneration. If soil macroporosity develops slowly (over pedogenic timescales), then land-surface disturbance from tillage would reduce or altogether eliminate the trends observed for natural surface or subsurface layers. By contrast, Fig. 2 shows significant ($P < 0.001$) relationships with the residual effective porosity of ploughed surface layers and the climatological variables, as well as a reduction in variability around the trend line, suggesting that macroporosity may develop rapidly at the surface. It is likely that multiple mechanisms act in concert on surface layers to produce this macroporosity; nonetheless, these processes appear to form macroporosity within years to decades[12] and are therefore relevant for examining decadal-to-century-scale feedbacks between soil hydraulic properties, soil moisture and climate.

We calculated changes in saturated hydraulic conductivities, K_s, from current and predicted effective porosities using a generalized Kozeny–Carman equation[13] and end-of-century predictions of mean annual precipitation in the Coupled Model Intercomparison Project phase 5[19] assuming a representative concentration pathway 6 (CMIP5 RCP6) scenario for five regions in the USA that have been predicted to experience considerable changes in mean annual precipitation (Fig. 3). Our results show a wide range of expected changes to the magnitude of K_s,

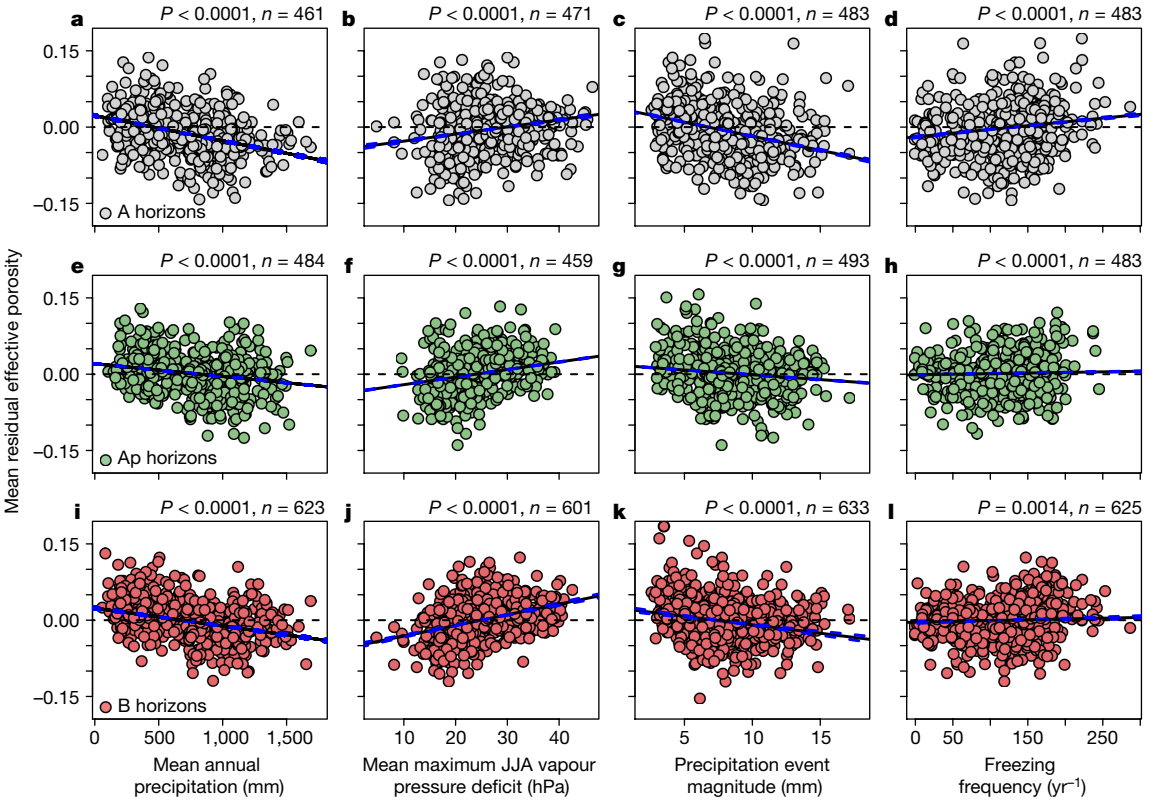


Fig. 2 | Climatological trends in mean residual effective porosity for natural surface, ploughed surface and subsurface layers. a–l, The mean residual effective porosity is negatively correlated with the mean annual precipitation (**a**, **e**, **i**) and the mean precipitation event magnitude (**c**, **g**, **k**), and positively correlated with the mean maximum vapour pressure deficit for June, July and August (JJA; **b**, **f**, **j**). The freezing frequency (**d**, **h**, **l**; in number of events per year) showed noticeable correlation with the mean residual effective porosity in the natural-surface layer only (**d**). Solid black (linear trends) and dashed blue (95% slope confidence intervals) lines are based on weighted linear regressions from 1,200 resampled data subsets. *n* indicates the size of the resampled data pool. Dashed black lines mark zero residual effective porosity for reference.

from −55% to +34%. With the exception of the Southeast Coastal Plain, mean predicted values were negative and ranged from −25% to −4%, signalling potentially less infiltration, more surface runoff and erosion and greater susceptibility to flash-flooding.

Overall, our findings provide observational evidence that macroporosity development is influenced by climate. This influence has not been previously considered in land–atmosphere forecasting, and it reinforces the hypothesis that climate change will probably intensify

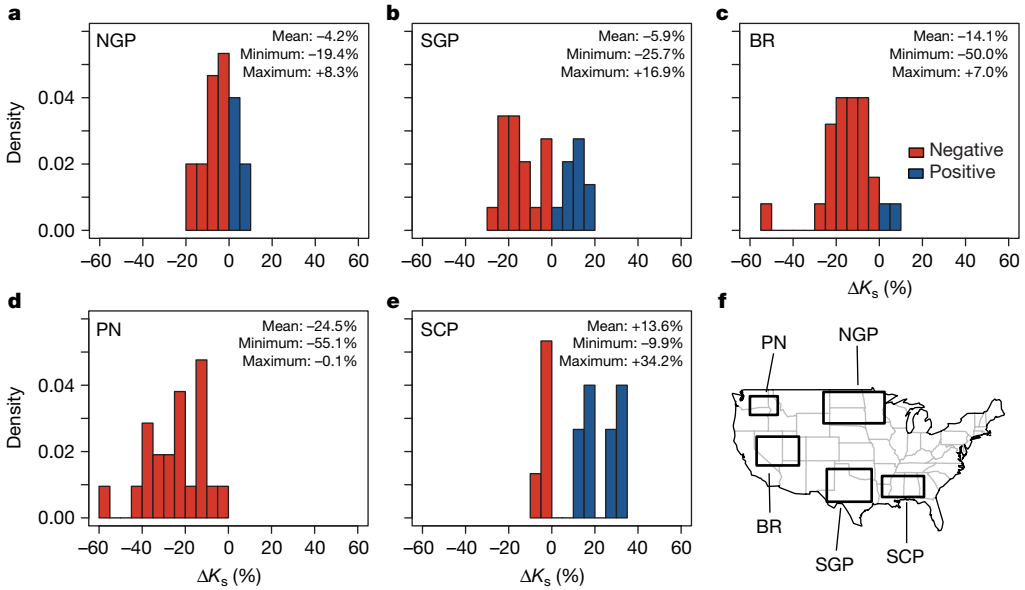


Fig. 3 | Expected per cent deviation of surface-layer (A horizon) saturated hydraulic conductivity from current values by the end of the century (2081–2100) for several regions in the USA. a–f, Regions presented are: Northern Great Plains (NGP; **a**), Southern Great Plains (SGP; **b**), Basin and Range (BR; **c**), Pacific Northwest (PN; **d**) and Southern Coastal Plain (SCP; **e**). The inset map (**f**) shows the locations of these regions. The largest mean magnitude changes are seen in PN, BR and SCP, in response to precipitation changes in these regions. The direction of the mean change is negative except in SCP.

the water cycle[20]. Increases in mean annual precipitation and event magnitude and decreases in atmospheric dryness and freezing temperature frequencies appear to minimize the expression of macroporosity. The rapid rate of macroporosity development for surface layers, along with its disproportionate effects on saturated hydraulic conductivity, may alter the distribution of soil moisture and affect related processes, such as evapotranspiration[21]. Our results suggest that these feedbacks should be incorporated into land–atmosphere parameterizations of regional and global climate models to better understand and predict the continental-scale hydrological cycle.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, statements of data availability and associated accession codes are available at https://doi.org/10.1038/s41586-018-0463-x.

Received: 11 July 2017; Accepted: 10 July 2018;
Published online 5 September 2018.

1. Beven, K. & Germann, P. Macropores and water flow in soils revisited. *Wat. Resour. Res.* **49**, 3071–3092 (2013).
2. Jarvis, N. J. A review of non-equilibrium water flow and solute transport in soil macropores: principles, controlling factors and consequences for water quality. *Eur. J. Soil Sci.* **58**, 523–546 (2007).
3. Janzen, H. H. et al. Global prospects rooted in soil science. *Soil Sci. Soc. Am. J.* **75**, 1–8 (2011).
4. Watson, K. W. & Luxmoore, R. J. Estimating macroporosity in a forest watershed by use of a tension infiltrometer. *Soil Sci. Soc. Am. J.* **50**, 578–582 (1986).
5. Lawrence, D. M. et al. Parameterization improvements and functional and structural advances in version 4 of the Community Land Model. *J. Adv. Model. Earth Syst.* **3**, M03001 (2011).
6. Clark, M. P. et al. Improving the representation of hydrologic processes in Earth System Models. *Wat. Resour. Res.* **51**, 5929–5956 (2015).
7. Hassink, J. The capacity of soils to preserve organic C and N by their association with clay and silt particles. *Plant Soil* **191**, 77–87 (1997).
8. Bronick, C. J. & Lal, R. Soil structure and management: a review. *Geoderma* **124**, 3–22 (2005).
9. Taina, I. A., Heck, R. J., Deen, W. & Ma, E. Y. T. Quantification of freeze–thaw related structure in cultivated topsoils using X-ray computer tomography. *Can. J. Soil Sci.* **93**, 533–553 (2013).
10. Brimhall, G. H. et al. Deformational mass transport and invasive processes in soil evolution. *Science* **255**, 695–702 (1992).
11. Platt, B. F., Kolb, D. J., Kunhardt, C. G., Milo, S. P. & New, L. G. Burrowing through the literature: the impact of soil-disturbing vertebrates on physical and chemical properties of soil. *Soil Sci.* **181**, 175–191 (2016).
12. Robinson, D. A. et al. Experimental evidence for drought induced alternative stable states of soil moisture. *Sci. Rep.* **6**, 20018 (2016).
13. Rawls, W. J., Giménez, D. & Grossman, R. Use of soil texture, bulk density, and slope of the water retention curve to predict saturated hydraulic conductivity. *Trans. ASAE* **41**, 983–988 (1998).
14. Nemes, A., Rawls, W. J. & Pachepsky, Y. A. Influence of organic matter on the estimation of saturated hydraulic conductivity. *Soil Sci. Soc. Am. J.* **69**, 1330–1337 (2005).
15. Tilman, D. Global environmental impacts of agricultural expansion: the need for sustainable and efficient practices. *Proc. Natl Acad. Sci. USA* **96**, 5995–6000 (1999).
16. Foley, J. A. et al. Global consequences of land use. *Science* **309**, 570–574 (2005).
17. West, L. T., Shaw, J. N. & Mersiovsky, E. P. in *The Soils of the USA* (eds West, L. T. et al.) Ch. 13 (Springer, Cham, 2017).
18. Peel, M. C., Finlayson, B. L. & McMahon, T. A. Updated world map of the Köppen–Geiger climate classification. *Hydrol. Earth Syst. Sci.* **11**, 1633–1644 (2007).
19. Taylor, K. E., Stouffer, R. J. & Meehl, G. A. An overview of CMIP5 and the experimental design. *Bull. Am. Meteorol. Soc.* **93**, 485–498 (2012).
20. Huntington, T. G. Evidence for intensification of the global water cycle: review and synthesis. *J. Hydrol.* **319**, 83–95 (2006).
21. Jung, M. et al. Recent decline in the global land evapotranspiration trend due to limited moisture supply. *Nature* **467**, 951–954 (2010).

Acknowledgements D.R.H. and D.G. thank R. Miskewitz for assistance in assigning Köppen–Geiger classes to the samples in the dataset. A.N., D.G. and D.R.H. thank the Norwegian Institute of Bioeconomy Research (NIBIO) for financial support. N.A.B. acknowledges funding support through USDA-AFRI 2014-67003-22070.

Reviewer information *Nature* thanks P. Hallett, D. Robinson and the other anonymous reviewer(s) for their contribution to the peer review of this work.

Author contributions D.R.H., D.G., A.N. and N.A.B. designed the study examining effective porosity with climate. A.N. and D.R.H. compiled the USDA-NRCS NCSS data. N.A.B. and C.J.W. compiled and analysed the atmospheric data. D.R.H. wrote the first draft of the paper and, with R.K., conducted the statistical analyses. All authors edited and commented on the manuscript and contributed to later iterations.

Competing interests The authors declare no competing interests.

Additional information
Extended data is available for this paper at https://doi.org/10.1038/s41586-018-0463-x.
Reprints and permissions information is available at http://www.nature.com/reprints.
Correspondence and requests for materials should be addressed to D.R.H.
Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

METHODS

Soil data. Soil information was selected from samples measured for sand (50–2,000 μm), silt (2–50 μm) and clay (less than 2 μm) contents, organic carbon content, field-capacity bulk density, air-dried bulk density and gravimetric water content at field capacity (−33 kPa) in the USDA-NRCS NCSS Characterization Database (https://ncsslabdatamart.sc.egov.usda.gov). In addition, only samples with recorded sampling locations were chosen. Samples representing surface layers were chosen if the midpoint of the soil morphological horizon (that is, the sampling interval) was within 25 cm of the land surface; subsurface layers were taken as those samples with midpoint depths between 25 cm and 100 cm of the land surface. A summary of the number of samples, names of the morphological horizons considered and depths are shown in Extended Data Table 1. We are unaware of any other soil database of this methodological consistency, extent and sample density, with soil physical properties measured directly as opposed to predicted via pedotransfer functions. Because this database contains relatively few samples in other areas of the world, we have restricted our analysis to within the USA.

Calculation of effective porosity (EP) and residual EP (REP). Total porosity (ϕ) was calculated using the bulk density of samples equilibrated to a pressure potential of $\rho_{fc} = -33$ kPa:

$$\phi = 1 - \frac{\rho_{fc}}{\rho_p}$$

where ρ_p is the average density of the soil particles, which is frequently assumed to be 2.65 Mg m^{-3}. The bulk density was also used to convert the gravimetric soil water content (θ_g) of samples equilibrated to −33 kPa to volumetric water content (θ_{fc}):

$$\theta_{fc} = \frac{\theta_g \rho_{fc}}{\rho_w}$$

where ρ_w is the density of water, assumed to be 1 Mg m^{-3}. Effective porosity was calculated as the difference between ϕ and θ_{fc} corresponding to pore sizes larger than 9 μm in diameter.

To account for possible effects that could arise from differences in particle-size distribution, soil organic matter and spatial location, EP was predicted ($\hat{EP}$) using a spatial error-regression model. The form of the regression equation linking EP to soil properties is

$$\begin{aligned}\hat{EP} = {} & C_0 + C_1 f_s + C_2 f_c + C_3 f_{SOC} + C_4 f_s f_c + C_5 f_c f_{SOC} \\ & + C_6 f_s f_{SOC} + C_7 f_s f_c f_{SOC} + \varepsilon\end{aligned} \quad (1)$$

where f_s, f_c and f_{SOC} are the mass fractions of sand, clay and soil organic carbon, respectively, C_i is the ith coefficient of the multiple regression model and ε denotes spatially correlated errors between neighbouring observations[22]. We used equation (1) to account for both direct effects on EP from sand, clay and soil organic carbon and effects that arise from their interactions for surface (A horizons) and subsurface (B horizons) layers separately.

Equation (1) is a multiple regression model, which can be represented using standard notation (simplified to one independent variable) as

$$y = X\beta + \varepsilon \quad (2)$$

where y is the dependent variable, β is the coefficient for the explanatory or independent variable X and ε represents an error vector. The form of the regression used had a spatial autoregressive process in the error term ε as follows[23]:

$$\varepsilon = \lambda W\varepsilon + \xi$$

Equation (2) can then be re-expressed as

$$y = X\beta + (I - \lambda W)^{-1}\xi$$

where λ is the spatial autoregressive coefficient for the error lag, $W\varepsilon$, W is a distance-based spatial weights matrix, ξ is an uncorrelated and homoscedastic error term, and I is the identity matrix. Thus, the error covariance becomes

$$\begin{aligned}E[\varepsilon\varepsilon'] &= \sigma^2 (I - \lambda W)^{-1}(I - \lambda W')^{-1} \\ &= \sigma^2 [(I - \lambda W)'(I - \lambda W)]^{-1}\end{aligned} \quad (3)$$

A spatial autoregressive error process leads to non-zero error covariances between every pair of observations. The error covariances decrease in magnitude with separation distance as the errors covary less owing to spatial autocorrelation when the separation distance is larger. The complex structure in the inverse matrices in equation (3) yields non-constant diagonal elements in the error covariance matrix, thus inducing heteroscedasticity in ε[23].

REP was calculated as the difference between the EP calculated from measurements of ρ_{fc} and θ_{fc} and $\hat{EP}$ calculated using equation (1):

$$REP = EP - \hat{EP} \quad (4)$$

REP expresses values of EP that are independent of particle size, soil organic carbon and spatial location and can therefore be used to investigate the effect of climate on EP.

Calculation of saturated hydraulic conductivity. The slope of the water retention curve (D) between θ_{fc} and the water content at the wilting point (θ_{wp}) was calculated as

$$D = \left| \frac{\log(\theta_{wp}/\theta_{fc})}{\log(1{,}500/33)} \right|$$

where 1,500 and 33 represent the absolute values of the pressure potentials at the wilting point and field capacity, respectively, in kilopascals[13]. The saturated hydraulic conductivity (in units of millimetres per hour) was calculated from D as

$$K_s = 1{,}930 EP^{3-D}$$

where 1,930 is an empirically determined coefficient[13].

Climatological data. Daily precipitation and minimum and maximum temperature data were obtained from all USHCN stations that had a continuous record between 1951 and 2011[24]. Precipitation data were used to calculate mean annual precipitation for each station. In addition, mean event magnitude and timing were fitted to precipitation data assuming exponential and Poisson distributions, respectively[25]. For each station, temperature data were used to calculate the freezing frequency as the mean number of times per year that the atmospheric temperature dropped below 0 °C. Mean maximum daily vapour pressure deficit values were obtained from the PRISM gridded climate data[26]. Predictions of end-of-century mean annual precipitation were obtained from the CMIP5[19] multi-model ensemble assuming a RCP6 scenario.

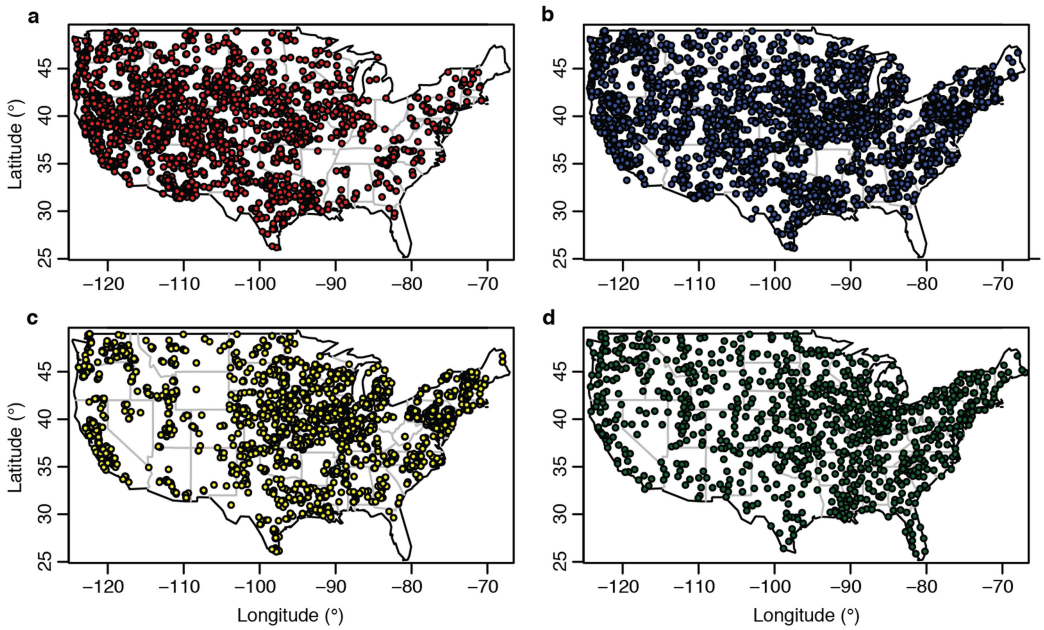
Statistical analysis. Effective porosity and climate point data were interpolated onto a 1° × 0.5° (approximately 85 km × 55 km) grid across continental USA using a tricubic-weighted least-squares quadratic trend surface[27] to smooth and visualize the spatial patterns in EP, REP and precipitation event magnitude and timing in Fig. 1 and Extended Data Fig. 2. Variograms of EP were calculated for each layer; REP values were calculated from equations (1) and (4) using a distance-based spatial-weight set based on the range parameter of the variograms to determine the spatially correlated error. The REP from soil samples with locations closest to each USHCN station were averaged and paired with the atmospheric data from each station. This aggregation procedure reintroduced some spatial autocorrelation; thus, we computed variograms to determine the scale of spatial variation in the aggregated data for each layer. We used a resampling method to randomly select subsamples of the aggregated data that were spaced at intervals that were at least twice the variogram range (600 km, 200 km and 840 km for the A, Ap and B horizons, respectively) to ensure spatial independence. This resampling method produced subsets of aggregated observations separated by distances greater than the intervals mentioned above. Weighted ordinary least-squares regression (that is, weighted by the number of samples averaged to represent each aggregated observation) was run on each subset and the presence of spatial autocorrelation in the residuals checked using Moran's I. When significant spatial auto-correlation was identified, that random subset was discarded and another selected until a total of 1,200 subsets was reached for each layer. Outliers in each subset were detected using an adjusted Mahalanobis distance[28] and removed prior to the regression analysis. Relationships between climatological variables and mean REP were examined using weighted ordinary least-squares regression for each of the 1,200 random subsets. The lines in Fig. 2 represent the average of these 1,200 regressions whereas the points represent the pool of samples from which the random subsets were selected. Spatial error regression was performed in GeoDa[22]. All other analyses were performed using R[29].

Code availability. The R scripts used to generate and analyse the data are publicly available in the GitHub repository, https://github.com/danielhirmas/nature2017-07-09186B.

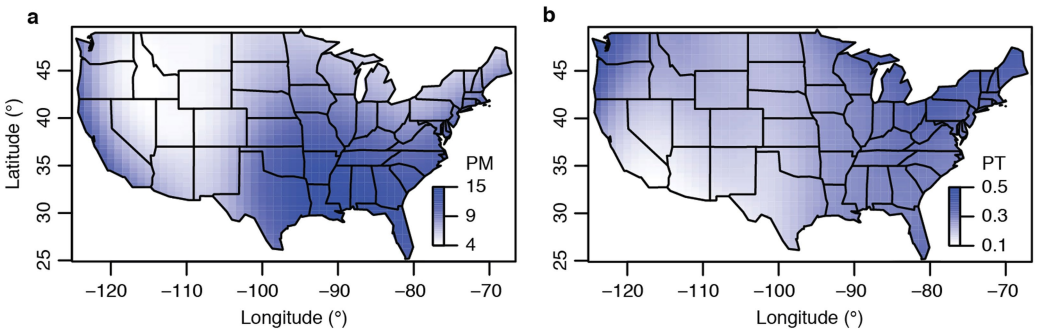
Data availability. The soil and climatological datasets generated and analysed during this study are publicly available in the GitHub repository, https://github.com/danielhirmas/nature2017-07-09186B. The soil datasets used in this study are also publicly available through the USDA-NRCS NCSS data repository, http://ncsslabdatamart.sc.egov.usda.gov/.

22. Anselin, L. *Exploring Spatial Data with GeoDa™: A Workbook* (Center for Spatially Integrated Social Science, 2005).
23. Anselin, L. & Bera, A. in *Handbook of Applied Economic Statistics* (eds Ullah, A. & Giles, D. E. A.) Ch. 7 (Marcel Dekker, New York, 1998).
24. Williams, C. N. Jr, Vose, R. S., Easterling, D. R. & Menne, M. J. *United States Historical Climatology Network Daily Temperature, Precipitation, and Snow Data.* (Oak Ridge National Laboratory, 2006).

25. Petrie, M. D. & Brunsell, N. A. The role of precipitation variability on the ecohydrology of grasslands. *Ecohydrology* **5**, 337–345 (2012).
26. Daly, C. et al. Physiographically sensitive mapping of climatological temperature and precipitation across the conterminous United States. *Int. J. Climatol*. **28**, 2031–2064 (2008).
27. Venables, W. N. & Ripley, B. D. *Modern Applied Statistics with S* 4th edn (Springer, New York, 2002).
28. Korkmaz, S., Goksuluk, D. & Zararsiz, G. MVN: An R package for assessing multivariate normality. *R J.* **6**, 151–162 (2014).
29. R Core Team. *R: A Language and Environment for Statistical Computing* version 3.0.2 (R Foundation for Statistical Computing, 2013).



Extended Data Fig. 1 | Distribution of selected soil samples and USHCN weather stations used in this study. a–d, Locations of A horizons (**a**), B horizons (**b**), Ap horizons (**c**) and USHCN weather stations (**d**). Weather stations that recorded continuous data since at least 1951 were selected. Soil sample data were obtained from the NCSS Characterization Database on 10 July 2013. Depth and soil morphological criteria for selection of the A, Ap and B horizon samples are given in Extended Data Table 1.



Extended Data Fig. 2 | Interpolated maps of mean precipitation magnitude per event and mean precipitation event timing from USHCN weather station data. a, Interpolated map of the mean precipitation magnitude per event (PM; in millimetres), calculated assuming that the magnitude of precipitation events followed an exponential distribution. **b**, Interpolated map of the mean precipitation event timing (PT; in events per day), calculated from a Poisson distribution for days with a precipitation event. Weather stations that recorded continuous data since at least 1951 were selected.

Extended Data Table 1 | Data selection criteria for samples used in this study

Soil layer	Depth* (cm)	n	Fraction of total† (%)	Representative horizons‡
A horizon	0-25	3582	44.8	A, Ac, Ag, Ass
Ap horizon	0-25	3248	40.6	Ap, Apc, Apg, Apt, ABp
B horizon	25-100	5756	37.6	B, Bw, Bt, Bg, Bss, Bc, Btss, Btg, Btc, Btm, Bwg, Bwc, Bssg

*Only samples with a horizon midpoint depth within the specified depth intervals were selected.
†The fraction of the total is given as the percentage of all A or B horizons in the database with the necessary data (location, bulk density, water contents at pressure potentials of −33 kPa and −1,500 kPa, sand, clay and organic carbon), falling within the respective depth intervals and in the conterminous USA.
‡Morphological horizon designations were selected to reduce effects from local addition of material (for example, carbonate) to the surface, from disturbance by management practices and from lithologic discontinuities. Symbols represent: c, concretions or nodules; g, strong gley colours; m, continuous pedogenic cementation; p, ploughed layer; ss, slickensides; t, illuvial accumulation of silicate clay; and w, weak colour or structure within the B horizon.

LETTER

https://doi.org/10.1038/s41586-018-0441-3

Jurassic stem-mammal perinates and the origin of mammalian reproduction and growth

Eva A. Hoffman[1]* & Timothy B. Rowe[1,2,3]

Transformations in morphology, physiology and behaviour along the mammalian stem lineage were accompanied by profound modifications to reproduction and growth, including the emergence of a reproductive strategy characterized by high maternal investment in a small number of offspring[1,2] and heterochronic changes in early cranial development associated with the enlargement of the brain[3]. Because direct fossil evidence of these transitions is lacking, the timing and sequence of these modifications are unknown. Here we present what is, to our knowledge, the first fossil record of pre- or near-hatching young of any non-mammalian synapsid. A large clutch of well-preserved perinates of the tritylodontid *Kayentatherium wellesi* (Cynodontia, Mammaliamorpha) was found with a presumed maternal skeleton in Early Jurassic sediments of the Kayenta Formation. The single clutch comprises at least 38 individuals, well outside the range of litter sizes documented in extant mammals. This discovery confirms that production of high numbers of offspring represents the ancestral condition for amniotes, and also constrains the timing of a reduction in clutch size along the mammalian stem. Although tiny, the perinates have an overall skull shape that is similar to that of adults, with no allometric lengthening of the face during ontogeny. The only positive allometries are associated with the bones that support the masticatory musculature. *Kayentatherium* diverged just before a hypothesized pulse of brain expansion that reorganized cranial architecture at the base of Mammaliaformes[4–6]. The association of a high number of offspring and largely isometric cranial growth in *Kayentatherium* is consistent with a scenario in which encephalization—and attendant shifts in metabolism and development[7,8]—drove later changes to mammalian reproduction.

The new specimens were discovered in the matrix underlying the partial skeleton of a *Kayentatherium* adult recovered from the Early Jurassic Kayenta Formation of northeastern Arizona. Remains of numerous tiny juveniles were found within a perimeter formed by the ribs, forelimb elements and vertebrae of the adult, which also preserves teeth and fragmentary jaws (Figs. 1, 2e–f, Extended Data Figs. 1, 2, Extended Data Table 1a, b, Supplementary Video 1). The young are represented by ten mostly complete skulls, as well as isolated jaws, teeth and postcranial elements (Fig. 2a–d, Extended Data Table 1a, c, Supplementary Videos 2, 3). The immature bones are thin and diaphanous, yet well-preserved. Distortion other than lateral compression is minimal, and positional relationships within the skull and among regions of the body are often maintained. Although three tritylodontid species are reported from the Kayenta Formation, dental features confirm the identification of adult and young as *K. wellesi*[9–11] (Fig. 2c–g, Supplementary Videos 4, 5).

Because almost no variation can be detected in size, tooth development or degree of ossification, the young are interpreted as belonging to a single clutch, presumably of the associated adult. Clutch size was estimated conservatively as one-half of the number of dentaries visible by microscope inspection or in micro-computed tomography scans (Supplementary Information); a partial dentary was counted if it preserved two erupted molariforms (Extended Data Table 1a, Supplementary Table 1). The minimum clutch size obtained by this method was 38, which is more than twice the average litter size of any mammal but is similar to average clutch size in crocodilians and a few squamates (Fig. 3, Supplementary Table 2). The discovery of a large clutch in a stem mammal provides material evidence that producing high numbers of offspring is the ancestral condition for amniotes, and that small litters represent a derived mammalian trait[1,2]. There is no sign of eggshell with the new specimens—although the shell, if it was similar to that of monotremes and lepidosaurs, would have been leathery and unlikely to fossilize[12,13]. The skull length of the *Kayentatherium* young is 1/20 that of a large adult (Museum of Comparative Zoology (MCZ), specimen number 8812), and 1/10 that of a small adult (MCZ 8811), which on the basis of dental widths was probably similar in size to the presumed mother (Fig. 2c–g, Extended Data Table 1d). The relative tininess of the individuals, the absence of discernible tooth wear (Fig. 2c, d) and the uniformity of ontogenetic stage (which indicates that there was little competition among the offspring for resources[14]) all suggest that the young were newly hatched—if not embryos. We refer to them neutrally as perinates.

Although aggregations of juvenile stem mammals have occasionally been described previously, the smallest young reported thus far

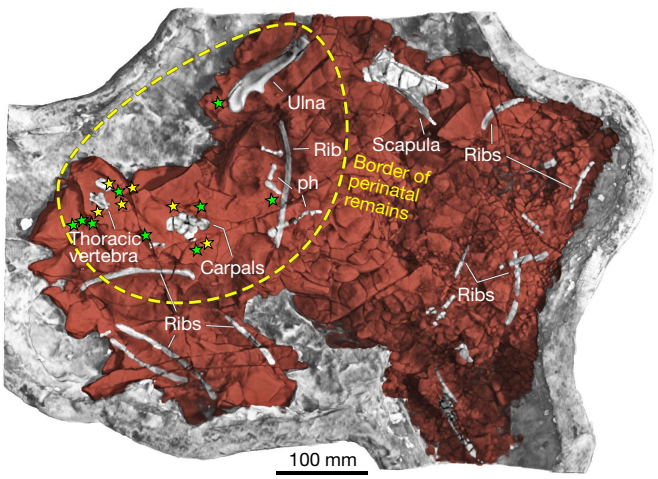


Fig. 1 | Volumetric rendering of original field jacket with excavated sediment containing adult and perinatal *Kayentatherium* remains (Vertebrate Paleontology Laboratory (TMM) 43690-5). See Supplementary Information for details on catalogue numbers and computed tomography scanning. Adult bones that had weathered to the surface are not shown. Stars mark the positions of selected perinates, illustrated in Extended Data Fig. 2. Yellow stars indicate perinatal individuals that preserve paired dentaries and that were counted towards our census; green stars indicate parts of perinatal individuals that were not counted towards our census (or counted as one-half; Extended Data Table 1a, Supplementary Table 1). Because stars are larger than perinatal bones, the positions of the stars are approximate. ph, phalanges.

[1]Jackson School of Geosciences, The University of Texas at Austin, Austin, TX, USA. [2]The University of Texas High-Resolution X-ray Computed Tomography Facility, The University of Texas at Austin, Austin, TX, USA. [3]Vertebrate Paleontology Laboratory, The University of Texas at Austin, Austin, TX, USA. *e-mail: eva.hoffman@utexas.edu

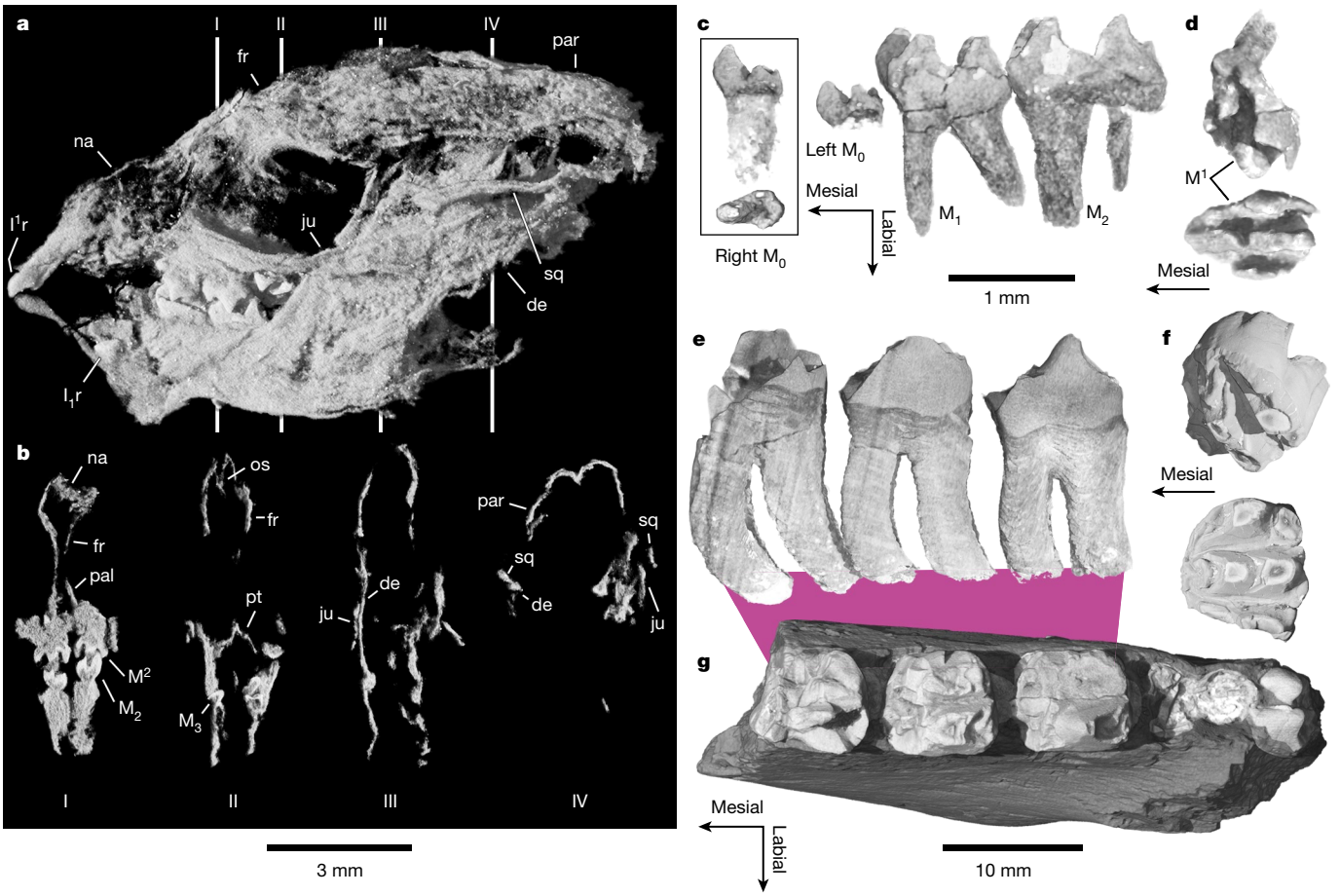


Fig. 2 | Skull of representative perinate, and dental anatomy of perinatal and adult *Kayentatherium*. **a**, **b**, Perinatal skull (TMM 43690-5.035a) and coronal sections at indicated levels (I–IV). The right side of the skull is shown reversed. For simplicity, the single upper and lower replacement incisors of *Kayentatherium* are designated I^1 and I_1, and their replacements I^1r and I_1r (but see previous studies[10,11]). **c**, **d**, Teeth of perinate (TMM 43690-5.017a). **c**, Right anteriormost lower molariform (M_0) in lingual and occlusal views, and left lower molariforms in labial view. **d**, Left upper molariform in labial (oblique) and occlusal views. **e**–**g**, Teeth of adult. **e**, Left lower molariforms in labial view. **f**, Upper molariform in labial (oblique) and occlusal views. **g**, Broken left dentary showing five lower molariforms in occlusal view. The distalmost molariform is unerupted. de, dentary; fr, frontal; I_1r, replacement lower incisor; I^1r, replacement upper incisor; ju, jugal; M_0–M_3, lower molariforms 0–3; M^1, M^2, upper molariforms 1, 2; na, nasal; os, orbitosphenoid; pal, palatine; par, parietal; pt, pterygoid; sq, squamosal.

are greater than one-fourth of adult size, and therefore well beyond the perinatal period[15,16]. By contrast, the new specimens provide fossil evidence regarding very early development in non-mammalian synapsids. Ontogenetic series of *Sphenodon punctatus*[17,18] (Fig. 4a–c) and *Monodelphis domestica* (Fig. 4d–g) are shown for comparison with *Kayentatherium* (Fig. 4h–j). In absolute size, the *Kayentatherium* perinates are similar to *Sphenodon* hatchlings and early *Monodelphis* pouch young (postnatal day 27; Fig. 4k).

The perinatal and adult skulls of *Kayentatherium* are notably similar in overall form. The growth of the face with respect to the braincase in reptiles and mammals is illustrated in Fig. 4. Whereas crocodilians and birds have autapomorphic extensions of the face[19], lepidosaurs retain the primitive amniote condition in which the skull grows mainly in isometry, with relative facial length constant during ontogeny (Fig. 4a–c). In mammals, the early skull accommodates a relatively much larger brain[3], with the result that embryos (or marsupial pouch young) have inflated braincases and short faces. Later in ontogeny, the expansion of the ethmoid olfactory skeleton, secondary palate and dentition causes the mammalian face to lengthen markedly as a proportion of the skull[20,21] (Fig. 4d–g).

In *Kayentatherium* skull ontogeny is plesiomorphic, with a relative facial length (measured from the zygomatic root to the tip of the premaxilla) that remains unchanged from perinate to large adult (Fig. 4h–j). As a formal test of isometry, we performed a standard-major-axis regression of facial length on skull length, a common proxy for overall body size[22]. The coefficient of allometry, *b*, equals 1.00, indicating that growth of the face is isometric (Extended Data Fig. 3, Extended Data Table 1d).

As in reptile ontogeny, positive allometries in *Kayentatherium* ontogeny mostly involve maturation of the masticatory system. During growth, the zygoma deepens and bows outward, a tall sagittal crest develops, the pterygoid transverse process lengthens and widens, and muscle attachments to the dentary expand and subtly reshape the coronoid process (Extended Data Fig. 4). These changes all reflect mature attainment of large, forceful masticatory muscles[23]. In our dataset, the maximum height of the zygomatic arch, length of the transverse process and width of the transverse process increase with positive allometry (coefficient of allometry or slope of log–log plot >1). Other skull measurements show isometry (Extended Data Fig. 3, Extended Data Table 1d).

The endocranial spaces and delicate bones of the palate, braincase and skull roof are visible in micro-computed tomography cross-sections of a perinatal skull (Fig. 2a, b, Supplementary Videos 6–8). Inferred brain shape reflects both phylogeny and ontogeny. The orbitosphenoids are high, as in other non-mammaliaform cynodonts, which indicates that the anterior forebrain is small[6]. However, the parietals preserve impressions of cerebral hemispheres that are tall domes separated by a deep interhemispheric sulcus (Fig. 2b), a condition that is not observed in more basal cynodonts. In the *Kayentatherium* perinates, the brain is wide relative to the rest of the skull, but later in development the brain becomes relatively narrow (Fig. 4h–j). Apart from the domed, separated cerebral hemispheres—which persist into adulthood—adult

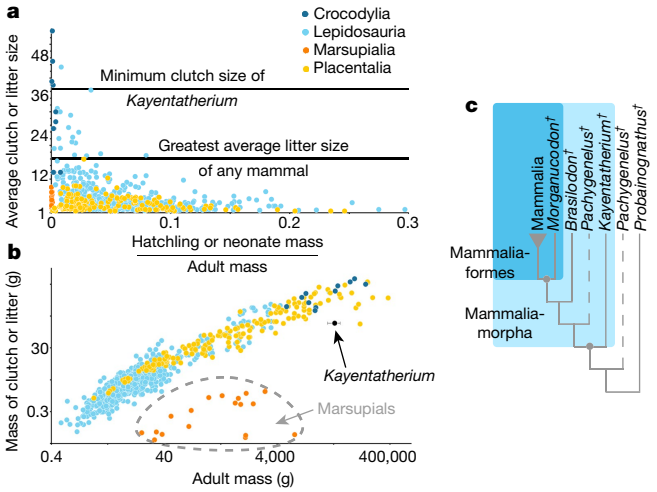


Fig. 3 | Clutch size of *Kayentatherium* and other amniotes, and phylogenetic position of *Kayentatherium*. **a**, Graph of average (mean or midpoint) clutch or litter size in 687 polytocous species of extant reptiles and mammals. The minimum size of the fossil clutch described here is indicated. **b**, Log–log plot of total clutch or litter mass versus adult mass (see Supplementary Information for explanation of body-mass estimation for adult and perinatal *Kayentatherium*). Data and sources can be found in Supplementary Table 2. **c**, Phylogeny of advanced cynodonts; relative phylogenetic positions of the included taxa are taken from previously published phylogenies[4,5]. Daggers indicate fossil taxa. Dotted lines indicate the two likeliest possibilities for the phylogenetic position of *Pachygenelus* with respect to *Kayentatherium*.

brain proportions are more similar to the tubular brains of more basal cynodonts than to the large, sub-spherical brains of mammals[6,24].

The teeth of the perinates are few in number and large relative to the jaws (Extended Data Fig. 4). Each complete jaw—upper and lower—contains an incisor and its replacement, two erupted molariforms and the cusps of two unerupted molariforms (Fig. 2a, Extended Data Fig. 5). The slender postdentary elements are distinguishable in the postdentary trough, and even the reflected lamina is thin and fragile but unbroken, with a pointed distal end (Extended Data Fig. 6).

Owing to their characteristic interlocking design, the teeth are commonly preserved in occlusion (Fig. 2a, b). Molariforms 1–4 are always present, and an additional, diminutive lower molariform is preserved in a single specimen on the left and right sides (Fig. 2c). No similar tooth has previously been reported in a tritylodontid. This tooth (M_0) is displaced on the right but preserved in situ on the left, anterior to the first pair of occluding molariforms. Its length is barely half that of the succeeding lower molariform (M_1) and its shape is unique, with three major cusps rather than four. The molariforms of tritylodontids are added in conveyor-belt fashion from the back of the tooth row, so that the anteriormost teeth are the oldest; teeth are eventually lost from the front of the jaw, opening a long diastema[11,25]. The small size and anomalous cusp configuration of M_0 suggest that it developed very early in ontogeny.

Distal to M_0, a strong mesiodistal size gradient is present along the tooth row and reflects the ongoing growth of the jaw[10,11,25] (Fig. 2c, Extended Data Fig. 5). If the rate of tooth formation was constant, the disproportionate size gap between M_0 and M_1 provides some evidence of faster growth earlier in development. On the basis of wear patterns, previous authors have proposed that tritylodontids had a delayed

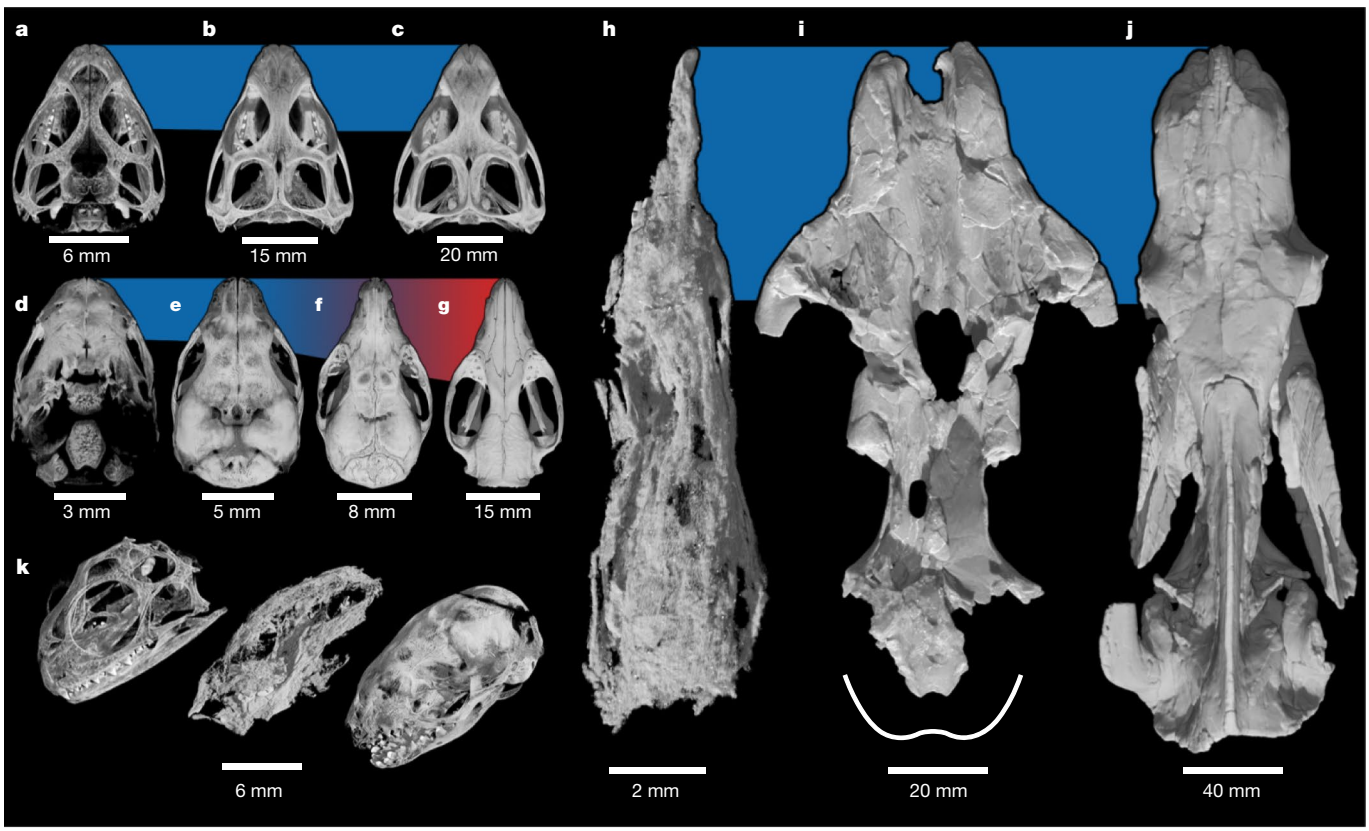


Fig. 4 | Early cranial ontogeny of reptiles, non-mammalian cynodonts and mammals. **a**–**j**, Ontogenetic series of *S. punctatus* (**a**–**c**; hatchling, Carnegie Museum 20660; small adult[17,18], Queen Mary University of London 0614; large adult, Yale Peabody Museum 9194), *M. domestica* (**d**–**g**; postnatal day 15, TMM M-7659; postnatal day 27, TMM M-8262; postnatal day 48, TMM M-7536; adult, TMM M-7599) and *K. wellesi* (**h**–**j**; perinate, TMM 43690-5.035a; small adult, MCZ 8811; large adult, MCZ 8812). The colour of the bar corresponds to the length of facial region, from the zygomatic root to the tip of the snout, relative to earliest ontogenetic stage. Skulls of TMM 43690-5.035a and MCZ 8812 are laterally compressed. **k**, *Sphenodon* hatchling, *Kayentatherium* perinate and *Monodelphis* pouch young (postnatal day 27), shown at equal magnification.

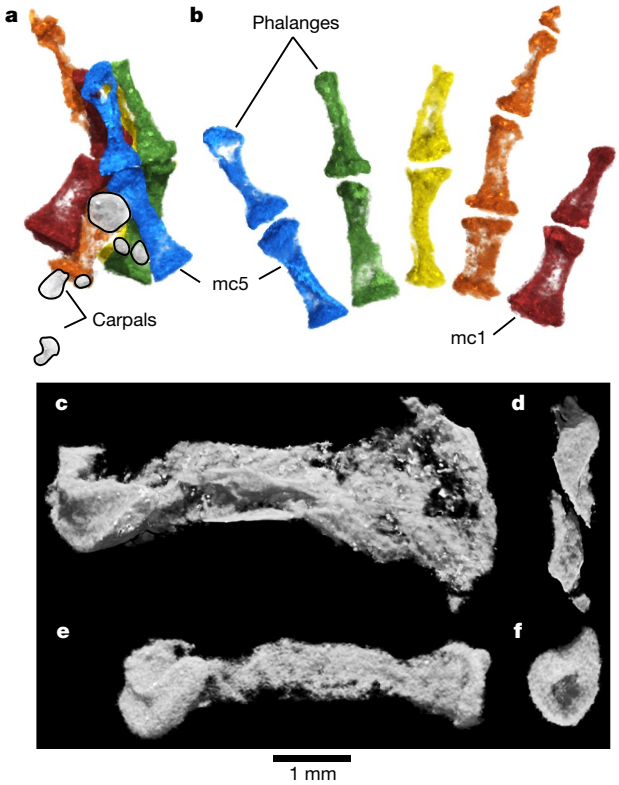


Fig. 5 | Manus and limb elements of *Kayentatherium* perinates. **a**, **b**, Partial left hand and wrist (TMM 43690-5.032a) in original position (**a**) and in dorsal view (**b**). **c**, **d**, Humerus (**c**; TMM 43690-5.032a) and its distal articular surface (**d**). **e**, **f**, Femur (**e**; TMM 43690-5.013a) and its distal articular surface (**f**). mc1, first metacarpal; mc5, fifth metacarpal.

onset of tooth development, with the implication that offspring must have required adult provisioning[26]. We show that tooth eruption in *Kayentatherium* began early in ontogeny. However, the single mesial cusp of M_0 seems almost certain to have inhibited occlusion, at least in the propalinal tritylodontid style[11,27]; therefore it appears unlikely that this early-generation molariform could have functioned in chewing.

In all the perinates, lower molariforms are better developed than upper molariforms. The first occluding lower molariform (M_1) always has two fully developed roots; the roots of M_2 are partially developed; and M_3 and M_4 are rootless, unconsolidated cusps. A similar pattern holds for the upper molariforms, although their roots are generally less distinct (Fig. 2c, d, Extended Data Fig. 5, Supplementary Videos 4, 5). In tritylodontids and haramiyids, successive tooth eruption—in conjunction with the rapid growth and elongation of the dentary—results in 'mesial drift' of the molariform crowns after root formation, so that over time the crowns are pushed forward and the implanted roots curve posteriorly[28,29] (Fig. 2e). Roots of the perinates are straight rather than curved, consistent with their immaturity. The anteriormost molariform (M_0) appears to have a single root (Fig. 2c), which on the left side either broke off post-mortem or was resorbed as the jaw grew and newer teeth moved forward. That the loss of M_0 was imminent, and its connection to the dentary tenuous, may explain the absence of that tooth in most of the specimens.

Postcrania of the perinates include elements of both forelimbs and hindlimbs. Examination of micro-computed tomography scans revealed a semi-articulated hand and wrist curled beneath the jaws of a partial skull (Fig. 5a, b, Extended Data Fig. 7). The carpals are ossified disks that have yet to take on mature shape and complex articulations. Their presence is notable, given that wrist elements tend to develop relatively late across tetrapods[30]. Notable, too, are the expanded proximal and distal ends and complex articular surfaces of the metacarpals and phalanges—features not developed in similar-sized mammals[31] or reptiles[32–34]. The limb bones of the perinates likewise exhibit a relatively mature shape. For example, the humerus has epicondyles and an incipient capitulum and ulnar condyle (Fig. 5c, d, Extended Data Fig. 8). These features reflect the epigenetic responsiveness of the developing bone to mechanical loading by muscle contraction, which begins before hatching[35]. The length of the perinatal femur is 83% that of the humerus (4.8 versus 5.8 mm), compared with a roughly even ratio in adults[36]. The presence of a reasonably large femur in the *Kayentatherium* perinates suggests that only a slight delay in hindlimb ossification, ancestral for tetrapods, is maintained on the mammalian stem lineage and by inference present primitively in mammals[31].

Limb elements in *Kayentatherium* may reflect a transitional phase in the histomorphological evolution of endochondral ossification, in association with changes in growth rate relative to basal therapsids[37] and the acquisition of numerous derived features of the postcranial skeleton that persist in extant mammals[4]. In basal tetrapods and reptiles (except lepidosaurs), perichondral ossification outpaces endochondral ossification as both advance from the shaft towards the ends of each limb bone[35]. Proximally and distally, perichondral bone forms around still-cartilaginous 'cones', which only later calcify and then ossify to form the endochondral articular ends. In mammals, endochondral ossification proceeds at the same rate as perichondral ossification. Dense cancellous endochondral bone is deposited at the proximal and distal ends simultaneously with the perichondral bone forming the surrounding collar; no cartilage cone ever develops. Secondary ossification centres subsequently appear in the cartilaginous epiphyses. In the perinates, both ends of the humerus and femur have short zones of cancellous bone that are denser than the perichondral shafts, indicating that endochondral ossification is relatively advanced—a condition approaching that of mammals (Fig. 5c–f, Extended Data Fig. 8). However, the ossified ends appear to have grown around cartilage cones, as in ancestral amniotes: the dense cancellous bone encloses conical indentations in the proximal end of the humerus and both ends of the femur (Fig. 5c–f, Extended Data Fig. 8). Secondary ossifications are absent. In mammals, endochondral ossification proceeds at an even greater rate, and cartilage cones no longer form, with secondary ossification centres taking their place[4,35].

The new specimens show that tritylodontids retained a primitive pattern of reproduction despite sharing a number of derived skeletal features with mammals[4,25,38]. The association of a plesiomorphic large clutch size and isometric cranial growth in a basal mammaliamorph suggests a link between reproductive strategy and brain size[1,2,7,8]. However, the first mammaliamorphs may have achieved more balanced rates of perichondral and endochondral ossification, in association with a postcranial skeleton that now resembled that of mammals more closely than that of basal cynodonts[4]. The origin of Mammaliaformes coincided with a 50% increase in relative brain size and emergence of the neocortex[6,39]. With the origin of crown Mammalia, further brain enlargement was unequivocally associated with reduced clutch size. The exceptional preservation of a large clutch of *Kayentatherium* perinates confirms a close correspondence among these features, and provides a more nuanced historical sequence in the evolution of mammalian characters.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, statements of data availability and associated accession codes are available at https://doi.org/10.1038/s41586-018-0441-3.

Received: 14 March 2018; Accepted: 17 July 2018;
Published online 29 August 2018.

1. Hopson, J. A. Endothermy, small size, and the origin of mammalian reproduction. *Am. Nat.* **107**, 446–452 (1973).
2. Case, T. J. On the evolution and significance of postnatal growth rates in vertebrates. *Q. Rev. Biol.* **53**, 243–282 (1978).
3. Koyabu, D. et al. Mammalian skull heterochrony reveals modular evolution and a link between cranial development and brain size. *Nat. Commun.* **5**, 3625 (2014).

4. Rowe, T. B. Definition, diagnosis, and origin of Mammalia. *J. Vert. Paleont.* **8**, 241–264 (1988).
5. Liu, J. & Olsen, P. The phylogenetic relationships of Eucynodontia (Amniota: Synapsida). *J. Mamm. Evol.* **17**, 151–176 (2010).
6. Rowe, T. B., Macrini, T. E. & Luo, Z.-X. Fossil evidence on origin of the mammalian brain. *Science* **332**, 955–957 (2011).
7. Sacher, G. A. & Staffeldt, E. F. Relation of gestation time to brain weight for placental mammals: implications for the theory of vertebrate growth. *Am. Nat.* **108**, 593–615 (1974).
8. Martin, R. D. Relative brain size and basal metabolic rate in terrestrial vertebrates. *Nature* **293**, 57–60 (1981).
9. Kermack, D. M. A new tritylodontid from the Kayenta Formation of Arizona. *Zool. J. Linn. Soc.* **76**, 1–17 (1982).
10. Sues, H.-D. First record of the tritylodontid *Oligokyphus* (Synapsida) from the Lower Jurassic of western North America. *J. Vert. Paleont.* **5**, 328–335 (1985).
11. Sues, H.-D. Skull and dentition of two tritylodontid synapsids from the Lower Jurassic of western North America. *Bull. Mus. Comp. Zool.* **151**, 217–268 (1986).
12. Hill, J. P. V. The development of the Monotremata. Part II. The structure of the egg-shell. *Trans. Zool. Soc. Lond.* **24**, 443–456 (1933).
13. Sander, P. M. Reproduction in early amniotes. *Science* **337**, 806–808 (2012).
14. Mock, D. W. & Parker, G. A. *The Evolution of Sibling Rivalry* (Oxford Univ. Press, Oxford, 1998).
15. Sánchez-Villagra, M. R. Developmental palaeontology in synapsids: the fossil record of ontogeny in mammals and their closest relatives. *Proc. R. Soc. Lond. B* **277**, 1139–1147 (2010).
16. Jasinoski, S. C. & Abdala, F. Aggregations and parental care in the Early Triassic basal cynodonts *Galesaurus planiceps* and *Thrinaxodon liorhinus*. *PeerJ* **5**, e2875 (2017).
17. Regnault, S., Hutchinson, J. R. & Jones, M. E. H. Sesamoid bones in tuatara (*Sphenodon punctatus*) investigated with X-ray microtomography, and implications for sesamoid evolution in Lepidosauria. *J. Morphol.* **278**, 62–72 (2017).
18. Regnault, S. & Hutchinson, J. R. Sesamoid bones in tuatara. *Open Science Framework* https://osf.io/bds35/ (2017).
19. Young, N. M. et al. Embryonic bauplans and the developmental origins of facial diversity and constraint. *Development* **141**, 1059–1063 (2014).
20. Wealthall, R. J. & Herring, S. W. Endochondral ossification of the mouse nasal septum. *Anat. Rec.* **288A**, 1163–1172 (2006).
21. Cardini, A. & Polly, P. D. Larger mammals have longer faces because of size-related constraints on skull form. *Nat. Commun.* **4**, 2458 (2013).
22. Jasinoski, S. C., Abdala, F. & Fernandez, V. Ontogeny of the Early Triassic cynodont *Thrinaxodon liorhinus* (Therapsida): cranial morphology. *Anat. Rec. (Hoboken)* **298**, 1440–1464 (2015).
23. Crompton, A. W. & Parker, P. Evolution of the mammalian masticatory apparatus. *Am. Sci.* **66**, 192–201 (1978).
24. Rowe, T. Coevolution of the mammalian middle ear and neocortex. *Science* **273**, 651–654 (1996).
25. Kuhne, W. G. *The Liassic therapsid Oligokyphus* (British Museum, London, 1956).
26. Hu, Y., Meng, J. & Clark, J. M. A new tritylodontid from the Upper Jurassic of Xinjiang, China. *Acta Palaeontol. Pol.* **54**, 385–391 (2009).
27. Crompton, A. W. Postcanine occlusion in cynodonts and tritylodontids. *Bull. Br. Mus.* **21**, 27–71 (1972).
28. Cui, G. & Sun, A. Postcanine root system in tritylodonts. *Vertebrata Palasiatica* **25**, 245–259 (1987).
29. Luo, Z.-X., Gatesy, S. M., Jenkins, F. A. Jr, Amaral, W. W. & Shubin, N. H. Mandibular and dental characteristics of Late Triassic mammaliaform *Haramiyavia* and their ramifications for basal mammal evolution. *Proc. Natl Acad. Sci. USA* **112**, E7101–E7109 (2015).
30. Richardson, M. K. et al. Heterochrony in limb evolution: developmental mechanisms and natural selection. *J. Exp. Biol.* **312B**, 639–664 (2009).
31. Weisbecker, V. Monotreme ossification sequences and the riddle of mammalian skeletal development. *Evolution* **65**, 1323–1335 (2011).
32. Rieppel, O. Studies on skeleton formation in reptiles. V. Patterns of ossification in the skeleton of *Alligator mississippiensis* Daudin (Reptilia, Crocodylia). *Zool. J. Linn. Soc.* **109**, 301–325 (1993).
33. Rieppel, O. Studies on skeleton formation in reptiles: patterns of ossification in the skeleton of *Chelydra serpentina* (Reptilia, Testudines). *J. Zool. (Lond.)* **231**, 487–509 (1993).
34. Rieppel, O. Studies on skeleton formation in reptiles. I. The postembryonic development of the skeleton in *Cyrtodactylus pubisulcus* (Reptilia: Gekkonidae). *J. Zool. (Lond.)* **227**, 87–100 (1992).
35. Carter, D. R., Mikić, B. & Padian, K. Epigenetic mechanical factors in the evolution of long bone epiphyses. *Zool. J. Linn. Soc.* **123**, 163–178 (1998).
36. Sues, H.-D. & Jenkins, F. A. in *Amniote Paleobiology: Perspectives on the Evolution of Mammals, Birds, and Reptiles* (eds Carrano, M. T. et al.) 114–152 (Chicago Univ. Press, Chicago, 2006).
37. Ray, S., Bandyopadhyay, S. & Bhawal, D. Growth patterns as deduced from bone microstructure of some selected neotherapsids with special emphasis on dicynodonts: phylogenetic implications. *Palaeoworld* **18**, 53–66 (2009).
38. Kemp, T. S. The relationships of mammals. *Zool. J. Linn. Soc.* **77**, 353–384 (1983).
39. Rowe, T. B. in *Evolution of Nervous Systems 2* Vol. 2 (ed. Kaas, J.) 1–52 (Academic, Oxford, 2017).

Acknowledgements We thank A. Zaman and B. Niessemeyer of the Navajo Nation Minerals Division for issuing the permit (dated 27 March 2000) under which this specimen was collected, and K. Calsoyas and T. Anderson for promoting our collaboration with the Navajo EcoScouts Program. Funding was provided by the National Science Foundation (EAR 1561622, IIS-9874781), by the Geology Foundation of The University of Texas and by the Jackson School of Geosciences. We thank B. Andres, S. Egberts, J. Franzosa, R. Gary, E. Gordon, T. Macrini, P. Owen, C. Sagebiel and R. S. Wallace for field, laboratory and curatorial assistance; M. Colbert, J. Maisano, J. Berlin and G. Rogers for computed tomography scanning the specimens described here; and S. Regnault, J. Hutchinson, C. Bell and digimorph.org for *Sphenodon* computed tomography scans.

Reviewer information *Nature* thanks H. Sues, L. Wilson and the other anonymous reviewer(s) for their contribution to the peer review of this work.

Author contributions T.B.R. directed collecting and preparation of the specimen. E.A.H. performed measurements, quantitative analyses and segmentation of computed tomography data, assembled comparative reproductive data and prepared the figures and tables. E.A.H. and T.B.R. wrote the manuscript and Supplementary Information.

Competing interests The authors declare no competing interests.

Additional information
Extended data is available for this paper at https://doi.org/10.1038/s41586-018-0441-3.
Supplementary information is available for this paper at https://doi.org/10.1038/s41586-018-0441-3.
Reprints and permissions information is available at http://www.nature.com/reprints.
Correspondence and requests for materials should be addressed to E.A.H.
Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

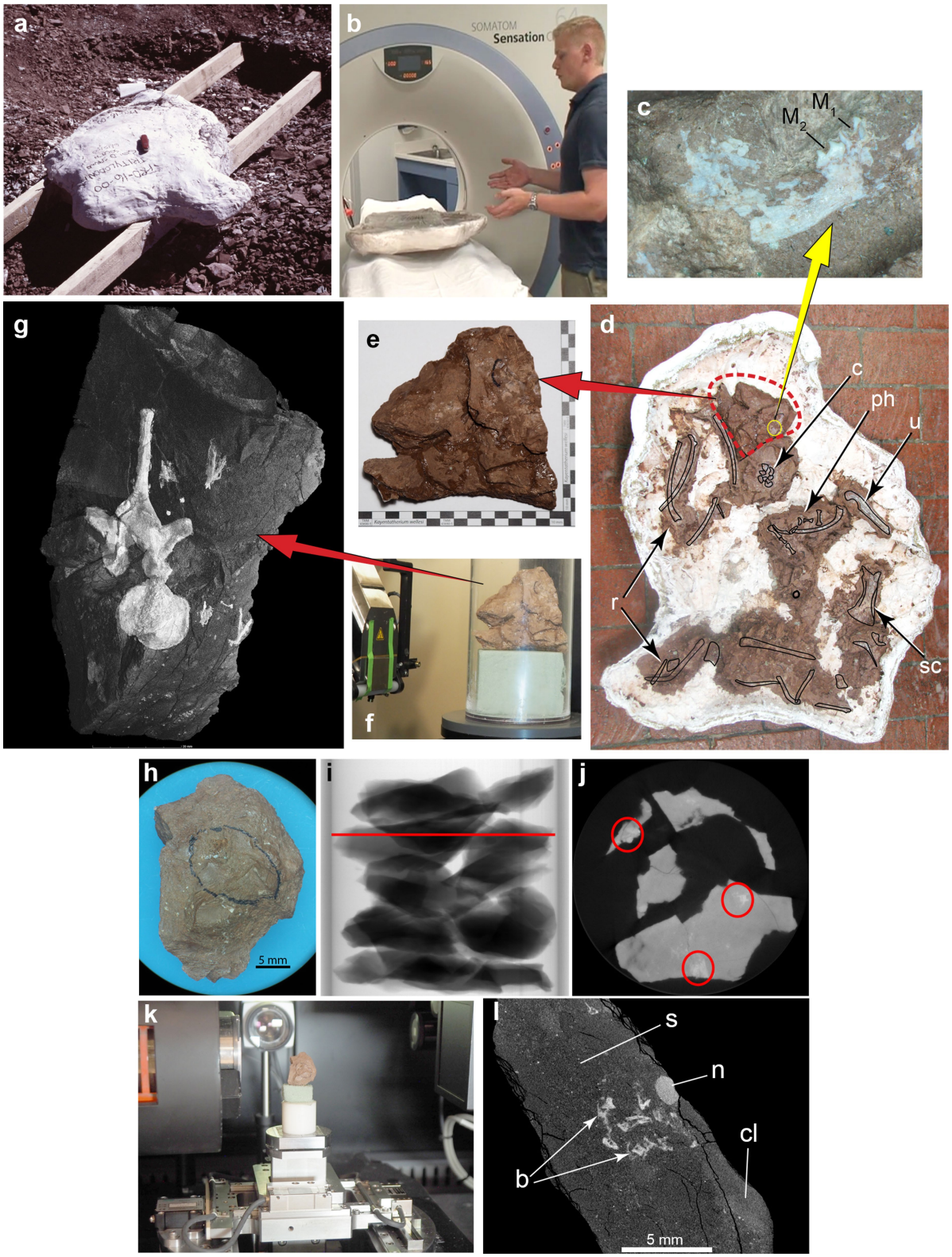
METHODS

Collection and mechanical preparation methods are described in the Supplementary Information. Initial computed tomography scans were performed at the Austin Heart Hospital, and other scan data reported here were generated by The University of Texas High-Resolution X-ray Computed Tomography Facility (UTCT). Detailed scan parameters are available in the Supplementary Information. Segmentation was performed in VGStudio Max 2.1. Measurements were performed by placing indicators on 3D isosurfaces in VGStudio Max 2.1. Standard-major-axis regressions were performed in R[40] using the package smatr[41]. Standard-major-axis regression was chosen for its appropriateness in cases involving uncertainty in the independent variable (here, skull length) as well as the dependent variable.

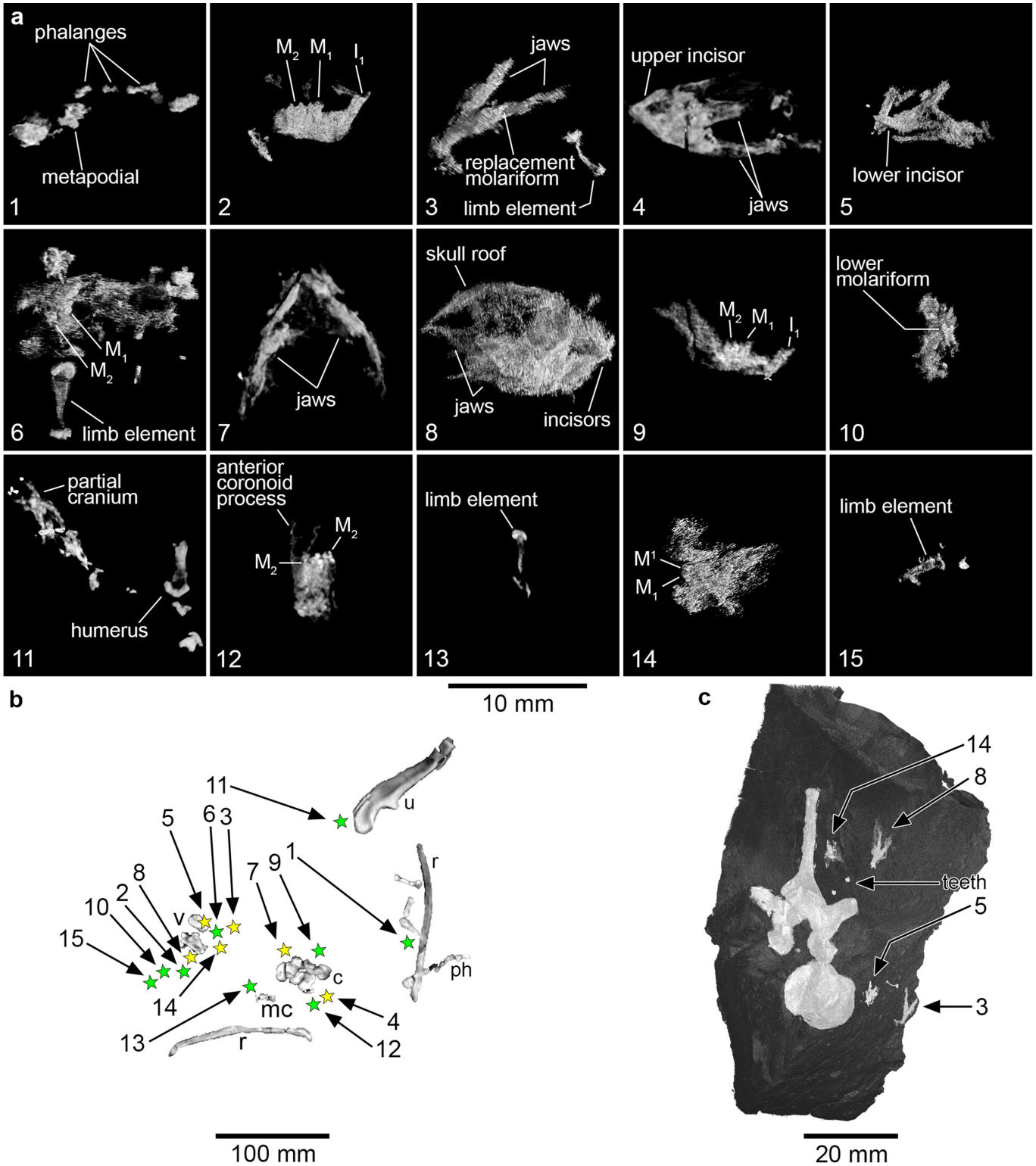
Reporting summary. Further information on experimental design is available in the Nature Research Reporting Summary linked to this paper.

Data availability. Computed tomography data are archived at The University of Texas High-Resolution X-ray Computed Tomography Facility (UTCT) and are available from the corresponding author upon reasonable request.

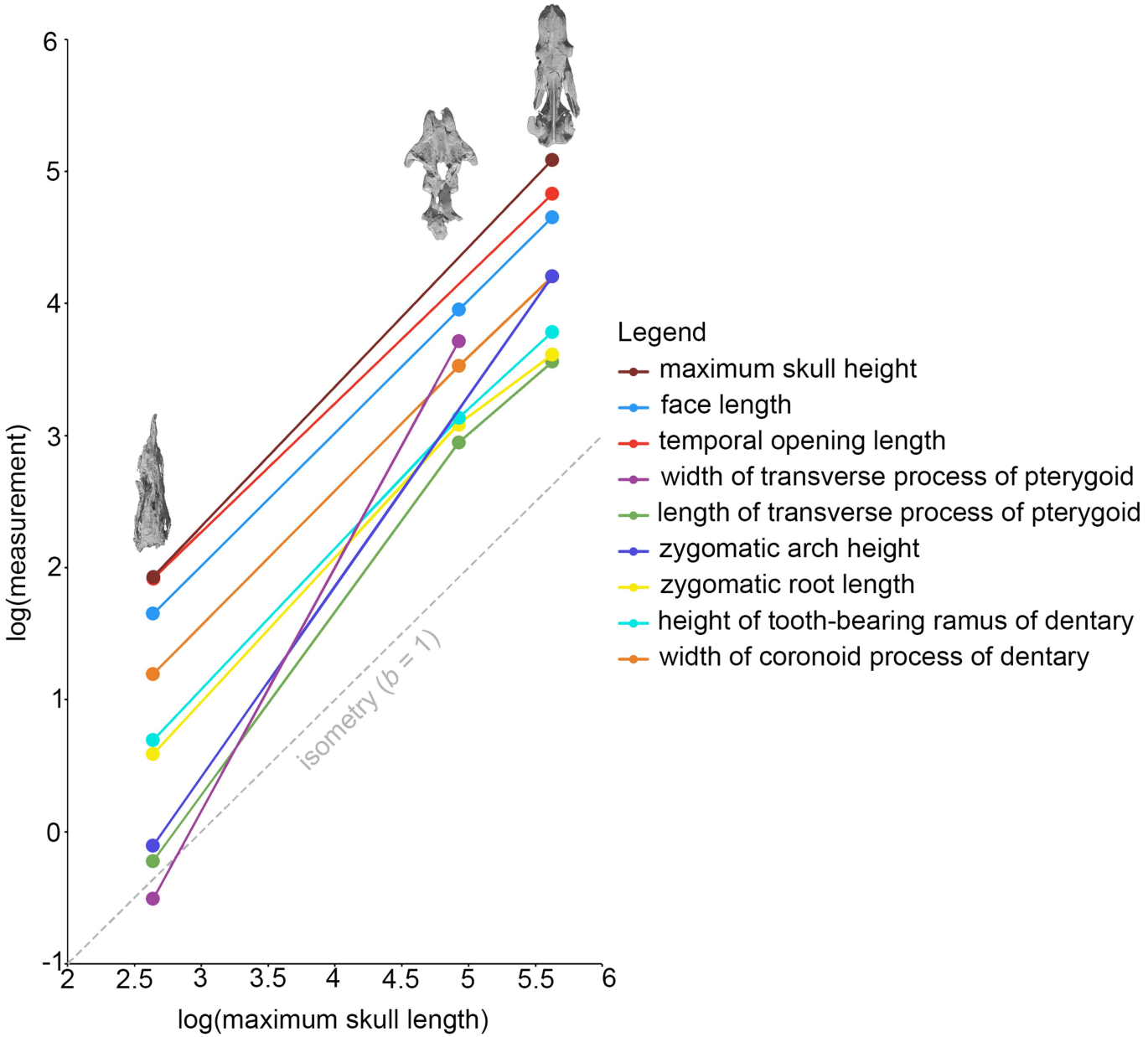
40. R Core Team. *R: A Language and Environment for Statistical Computing* https://www.R-project.org/ (R Foundation for Statistical Computing, Vienna, 2017).
41. Warton, D. I., Duursma, R. A., Falster, D. S. & Taskinen, S. smatr 3–an R package for estimation and inference about allometric lines. *Methods Ecol. Evol*. **3**, 257–259 (2012).



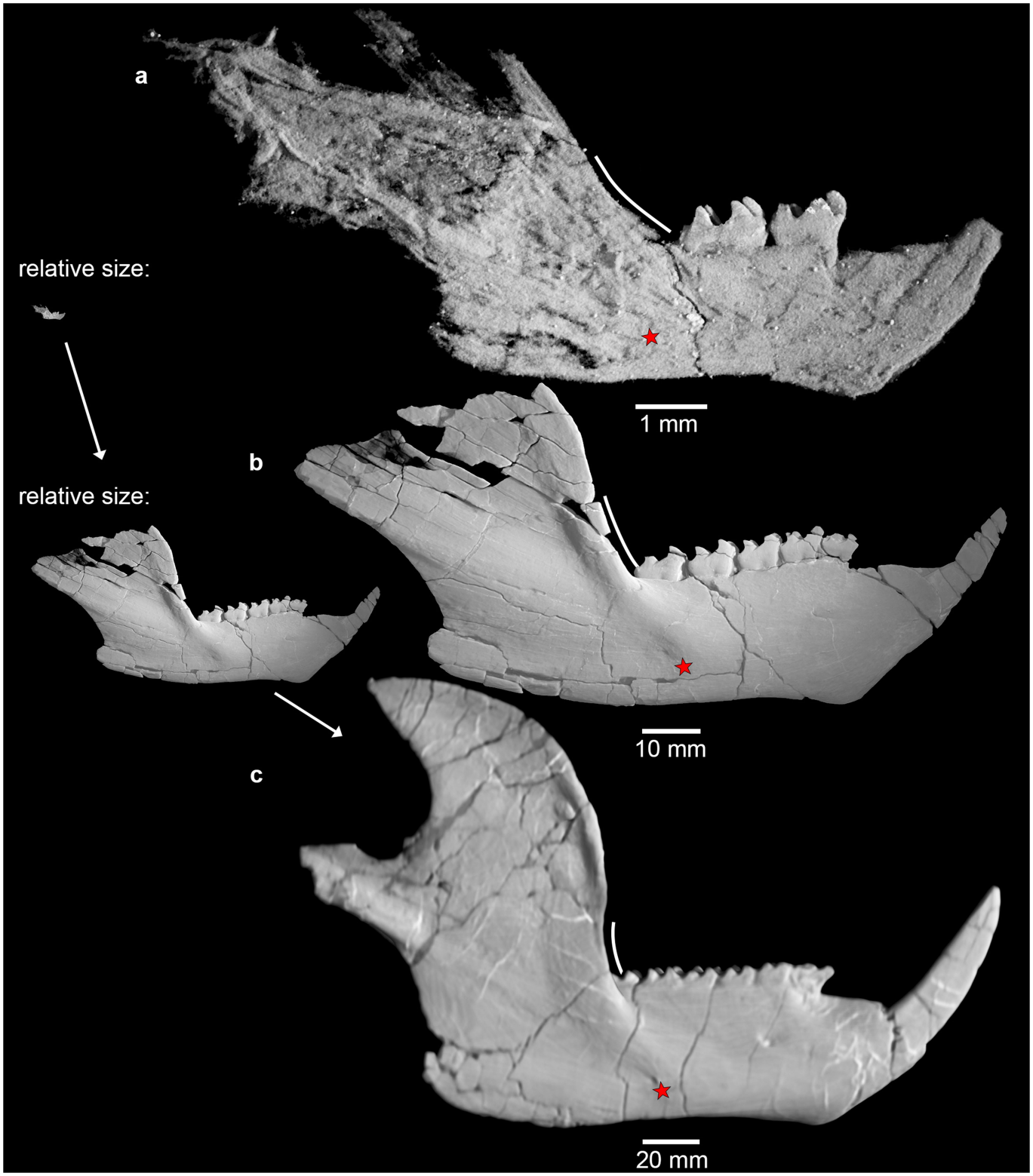
Extended Data Fig. 1 | Preparation and scanning of specimen (TMM 43690-5). a, Field jacket next to quarry, set on 8-foot-long, 2-by-4-inch beams; the red object on top is a Swiss army knife. **b**, Scanning the opened jacket at the Austin Heart Hospital. S. Egberts (right) discovered the perinates. **c**, **d**, Photomicrograph of a right mandible (TMM 43690-5.135d) (**c**) exposed on the surface of a 'chunk' of matrix removed from the opened jacket (**d**). At this stage of preparation, the jacket still contained maternal bones (black outline) as well as some perinatal remains. **e**, **f**, Chunk of matrix (TMM 43690-5.135) removed from jacket for high-resolution computed tomography scanning. **g**, Volumetric rendering of a sub-volume scan of the chunk showing a maternal thoracic vertebra surrounded by perinatal bones. **h**, Small chunk (TMM 43690-5.013) with flecks of perinatal bone exposed on the surface. **i**, **j**, Digital radiograph (**i**) of multiple flakes mounted in lucite tube for reconnaissance scan, and computed tomography slice (**j**), at level indicated by red line in **i**, showing perinatal remains (red circles). **k**, **l**, Computed tomography scanning (**k**) of individual chunk (TMM 43690-5.013), and high-resolution computed tomography slice (**l**) showing perinatal bones and components of the sediment. b, perinatal bones; c, carpals; cl, clay clast; M_1, M_2, lower molariforms 1, 2; n, carbonate nodule; ph, phalanges; r, ribs; s, sand matrix; sc, scapula; u, ulna.



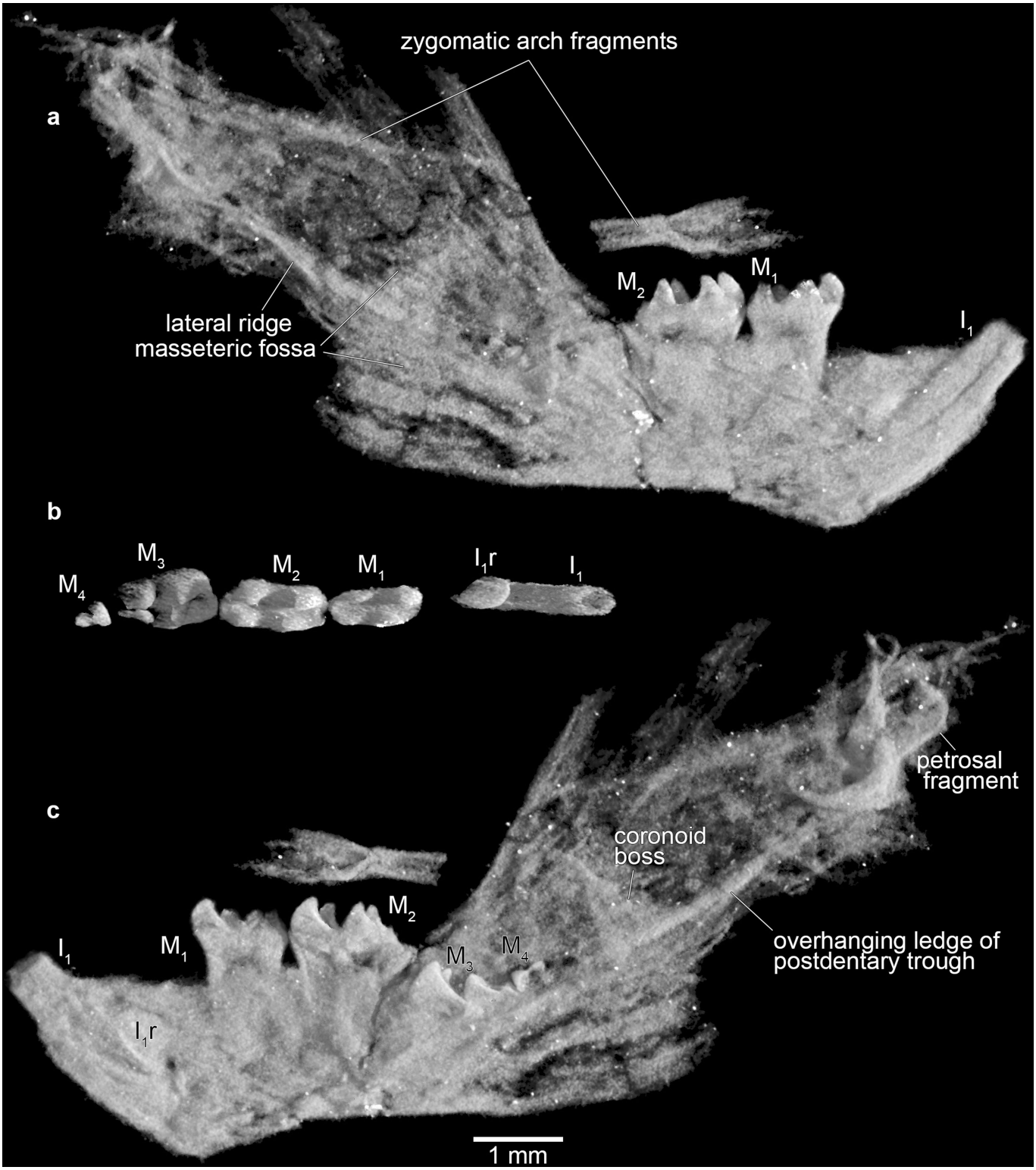
Extended Data Fig. 2 | Selected perinatal remains. a, Perinatal bones. **b**, Original positions of bones in **a** with respect to adult elements. Yellow stars indicate perinatal individuals that preserve paired dentaries and that were counted towards our census; green stars indicate parts of perinatal individuals that were not counted towards our census (or counted as one-half; Extended Data Table 1a, Supplementary Table 1). Because stars are larger than perinatal bones, the positions of the stars are approximate. **c**, Chunk of original matrix showing adult thoracic vertebra and selected perinatal remains in situ. Matrix is rendered transparent. The vertebra in **c** corresponds to that shown in **b**. Specimen numbers 2, 3, 5, 6, 8, 10, 14 and 15 correspond to TMM 43690-5.135a–TMM 43690-5.135h; specimen numbers 4, 7, 9, 12 and 13 correspond to TMM 43690-5.136a–TMM 43690-5.136e; specimen number 11 corresponds to TMM 43690-5.137a; and specimen number 1 corresponds to TMM 43690-5.139a. c, carpus; I_1, lower incisor; M_1, M_2, lower molariforms 1, 2; M^1, first upper molariform; mc, metacarpal; ph, phalanges; r, rib; u, ulna; v, vertebra.



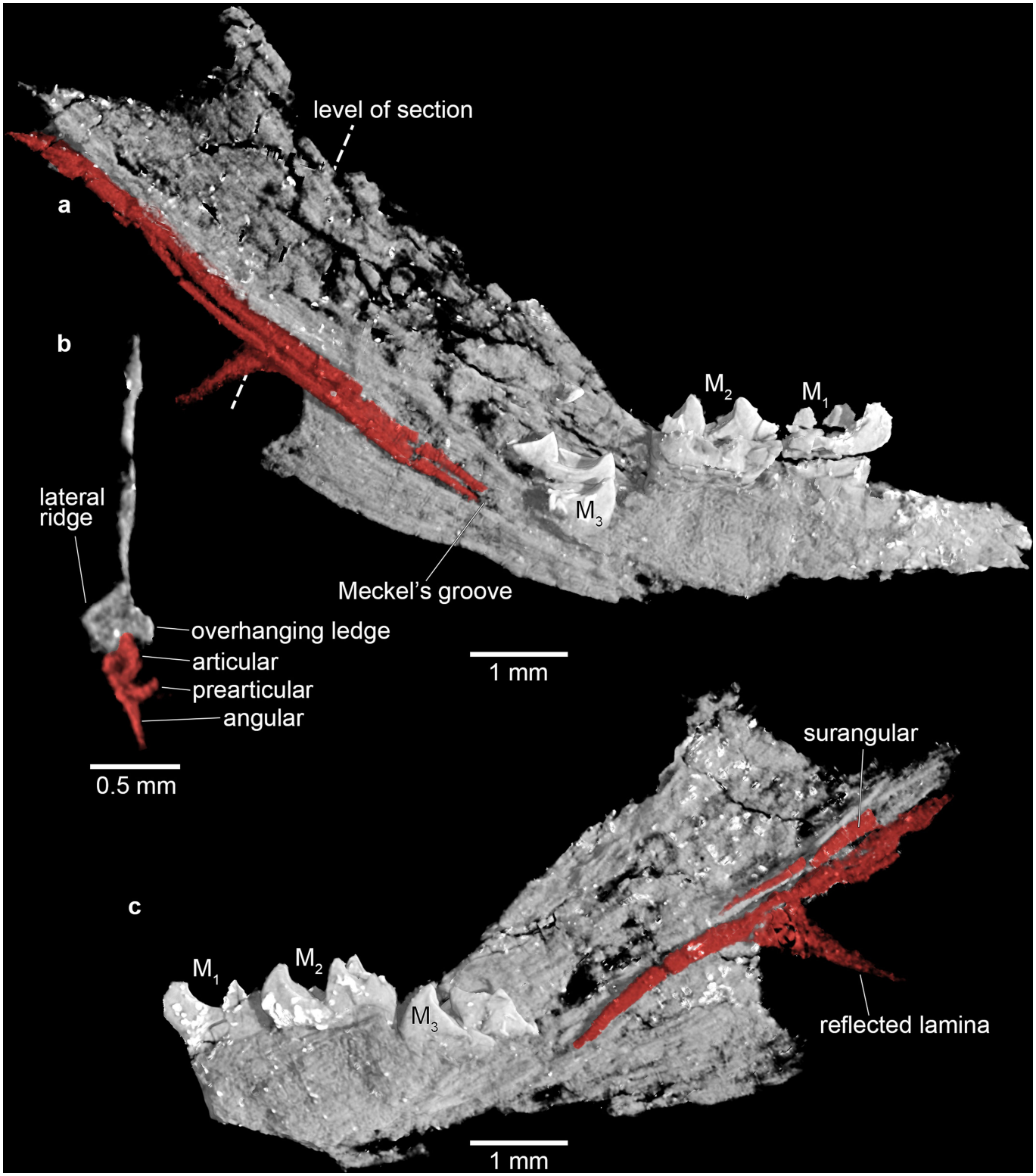
Extended Data Fig. 3 | Skull ontogeny in *Kayentatherium*. The log–log plot shows various skull measurements versus maximum skull length at up to three ontogenetic stages (perinate, small adult and large adult). Because sample sizes are low, lines connecting the data points are shown in place of regression lines. The reference line has a slope or coefficient of allometry, b, equal to 1. In analyses of allometry, $b < 1$ indicates negative allometry and $b > 1$ indicates positive allometry. Raw data are in Extended Data Table 1c. Skulls shown are TMM 43690-5.035a, MCZ 8811 and MCZ 8812.



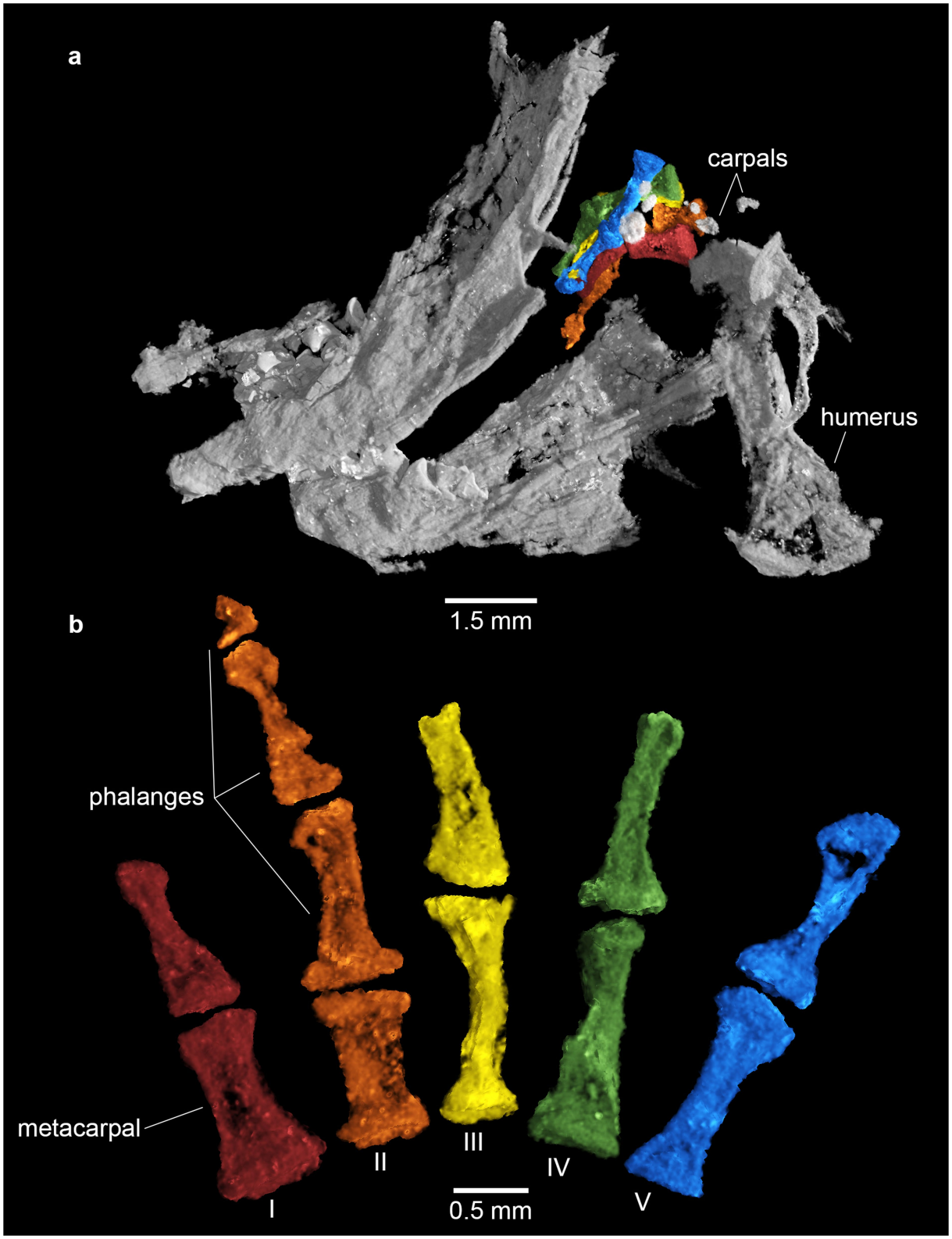
Extended Data Fig. 4 | Mandibular and dental ontogeny in *Kayentatherium*. **a**–**c**, Ontogenetic series of right *Kayentatherium* dentaries in lateral view (**a**, perinate, TMM 43690-5.035b; **b**, small adult, MCZ 8811; **c**, large adult, MCZ 8812). Curves follow the angle of the coronoid process at its base and stars indicate the anterior limit of the masseteric fossa. Jaws on the left are shown relative to the large adult, MCZ 8812.



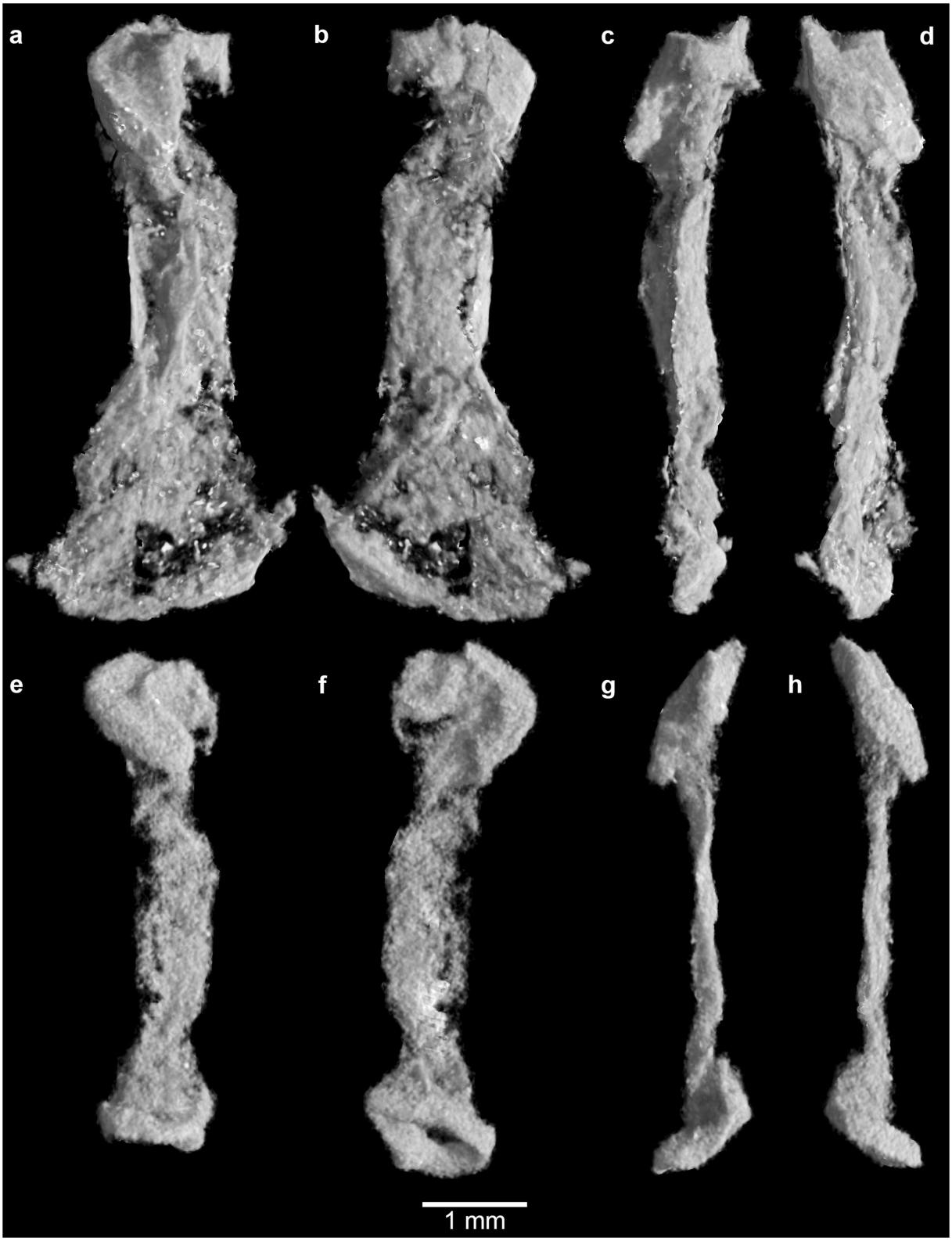
Extended Data Fig. 5 | Dental and mandibular anatomy of *Kayentatherium* perinate (TMM 43690-5.035b). **a**, Right dentary in lateral view. **b**, Teeth in dorsal view in situ without dentary. **c**, Right dentary in medial view. I_1, lower incisor; I_1r, replacement lower incisor; M_1–M_4, lower molariforms 1–4.



Extended Data Fig. 6 | Postdentary elements of *Kayentatherium* perinate (TMM 43690-5.032a). a, Partial left mandible in medial view, with postdentary elements coloured red. **b**, Section through coronoid process and postdentary elements at position indicated in **a**. **c**, Partial right mandible in medial view. M_1–M_3, lower molariforms 1–3.



Extended Data Fig. 7 | Additional views of hand and forelimb of *Kayentatherium* perinate (TMM 43690-5.032a). a, Partial left hand in situ with paired jaws, carpals, humerus and additional fragmentary elements. **b**, Partial left hand in palmar view. I–V, digits I–V.



Extended Data Fig. 8 | Additional views of humerus and femur of *Kayentatherium* perinates. a–d, Humerus (TMM 43690-5.032a) in frontal (**a**, **b**) and side (**c**, **d**) views. **e–h**, Femur (TMM 43690-5.013a) in frontal (**e**, **f**) and side (**g**, **h**) views.

Extended Data Table 1 | Sample of adult and perinatal elements, and skull measurements at three ontogenetic stages

a

Preserved part of skeleton (perinates)	Count	Minimum individuals
Pair of dentaries* in skull	10	10
Pair of dentaries* not in skull	10	10
Unpaired dentary* identified to side	4	2†
Unpaired dentary* not identified to side	32	16†
Total estimate		**38**

b

Preserved part of skeleton (adult)	Count	Surface float or excavated block
Maxillary fragments, containing a total of 6 molariforms	–	float
Isolated upper molariforms	4	float
Dentary, left, with 4 erupted molariforms and 1 unerupted molariform	1	float
Dentary, right, with incisor and 4 erupted molariforms	1	float
Humerus, right (broken)	1	float
Proximal ulna, right	1	float
Isolated metapodials (also see below)	2	float
Caudal vertebrae	3	float
Scapula, left	1	block
Ulna, left	1	block
Articulated carpals, left	7	block
Articulated digit I (1 metacarpal and 2 phalanges)	1	block
Isolated metacarpal	1	block
Isolated phalanges	2	block
Ribs and rib fragments	14	block
Thoracic vertebra	1	block

c

Preserved part of skeleton (perinates)	Count
Humeri	2
Articulated partial hand and wrist	1
carpals	6
metacarpals	5
phalanges	7
Femur	1
Isolated metapodial	1
Isolated phalanges	3
Unidentified limb elements	6

d

Skull measurement	Perinate‡ (mm)	Perinatal specimen	Small adult§ (mm)	Large adult‖ (mm)	*n*	R^2	*b*
Maximum skull length (premaxilla to parietal)	14.0	TMM 43690-5.035a	138 (est.)	278	3	–	–
Maximum skull height, including dentary	6.9	TMM 43690-5.035a	–	162	2	–	–
Face length (premaxilla to zygomatic root)	5.2	TMM 43690-5.035a	52	105	3	1.00	1.00
Temporal opening length	6.8	TMM 43690-5.035a	–	125	2	–	–
Width of transverse process of pterygoid	0.6	TMM 43690-5.045a	41	(distorted)	2	–	–
Length of transverse process of pterygoid	0.8	TMM 43690-5.045a	19	35	3	0.99	1.30
Zygomatic arch height	0.9	TMM 43690-5.035a	–	67	2	–	–
Zygomatic root length	1.8	TMM 43690-5.045a	22	37	3	1.00	1.03
Height of tooth-bearing ramus of dentary	2.0	TMM 43690-5.035b	23	44	3	1.00	1.04
Width (at base) of coronoid process of dentary	3.3	TMM 43690-5.035b	34	67	3	1.00	1.01

a, Summary of perinatal census. **b**, Inventory of adult elements. **c**, Inventory of identified perinatal postcrania. **d**, Skull measurements at three ontogenetic stages, with summary of regressions on skull length.

*Intact dentary or partial dentary including at least two (of three) erupted molariforms.

†As a conservative estimate, each unpaired dentary = 0.5 individuals.

‡All measurements were performed by placing indicators on 3D isosurfaces in VGStudio Max 2.1. Standard-major-axis regressions on skull length were performed in R[40] using the package smatr[41] for measurements available for $n = 3$ ontogenetic stages. R^2 is the square of the correlation coefficient. b is the coefficient of allometry, in which $b = 1$ indicates isometry, $b > 1$ indicates positive allometry and $b < 1$ indicates negative allometry.

§Measurements from MCZ 8811.

‖Measurements from MCZ 8812.

LETTER

https://doi.org/10.1038/s41586-018-0430-6

Sulfoxaflor exposure reduces bumblebee reproductive success

Harry Siviter[1]*, Mark J. F. Brown[1] & Ellouise Leadbeater[1]

Intensive agriculture currently relies on pesticides to maximize crop yield[1,2]. Neonicotinoids are the most widely used insecticides globally[3], but increasing evidence of negative impacts on important pollinators[4–9] and other non-target organisms[10] has led to legislative reassessment and created demand for the development of alternative products. Sulfoximine-based insecticides are the most likely successor[11], and are either licensed for use or under consideration for licensing in several worldwide markets[3], including within the European Union[12], where certain neonicotinoids (imidacloprid, clothianidin and thiamethoxam) are now banned from agricultural use outside of permanent greenhouse structures. There is an urgent need to pre-emptively evaluate the potential sub-lethal effects of sulfoximine-based pesticides on pollinators[11], because such effects are rarely detected by standard ecotoxicological assessments, but can have major impacts at larger ecological scales[13–15]. Here we show that chronic exposure to the sulfoximine-based insecticide sulfoxaflor, at dosages consistent with potential post-spray field exposure, has severe sub-lethal effects on bumblebee (*Bombus terrestris*) colonies. Field-based colonies that were exposed to sulfoxaflor during the early growth phase produced significantly fewer workers than unexposed controls, and ultimately produced fewer reproductive offspring. Differences between the life-history trajectories of treated and control colonies first became apparent when individuals exposed as larvae began to emerge, suggesting that direct or indirect effects on a small cohort may have cumulative long-term consequences for colony fitness. Our results caution against the use of sulfoximines as a direct replacement for neonicotinoids. To avoid continuing cycles of novel pesticide release and removal, with concomitant impacts on the environment, a broad evidence base needs to be assessed prior to the development of policy and regulation.

The widespread global use of highly effective neonicotinoid-based pesticides has led to the evolution of resistance among several insect crop pests[16] and has generated worldwide interest in emerging sulfoximine-based alternatives that have been shown to be effective in targeting some neonicotinoid-resistant species[17–19]. This potential lack of cross-resistance may reflect differences in the three-dimensional molecular structure that preclude the breakdown of sulfoximines by enzymes that are involved in neonicotinoid metabolism[20], supporting the claim that sulfoximines and neonicotinoids are chemically distinct[17]. However, as selective agonists of insect nicotinic acetylcholine receptors[17], the two pesticide groups share a common biological mode of action. This raises major concerns about potential effects on non-target species, and particularly on bees. Neonicotinoids, while not lethal to bees at field-realistic levels, have severe sub-lethal effects on both social and solitary bees, influencing cognition, foraging ability, homing ability, reproductive output, colony initiation[5,7,8,15,21–25], and, potentially, pollination services[26]. Mathematical modelling has shown that these sub-lethal stressors can have considerable negative consequences for colony fitness downstream in the colony cycle[14,15].

To assess whether sulfoxaflor, the first marketed sulfoximine-based pesticide, has similar negative effects on bees, we fed either untreated sucrose solution (1.8 M) or a sucrose solution containing 5 μg dm^{-3} (5 ppb) of sulfoxaflor to nascent *Bombus terrestris* colonies reared from wild-caught queens. We based this concentration on available estimates for sulfoxaflor residues in forager-collected nectar post-spray[27] (Extended Data Fig. 1a), because spray application is currently the most common application procedure (although products containing sulfoxaflor have also been developed for seed treatments and are already available for use on bee-pollinated crops in some markets[28]). After two weeks of laboratory-based exposure, size-matched colonies were placed in the field around a university parkland campus following a paired design and were no longer provided with additional resources. Staggered weekly nocturnal censuses revealed a clear difference in colony demographics between control and experimental colonies. The bumblebee colony cycle is characterized by an early growth phase in which worker numbers increase rapidly to create a large workforce, followed by a switch to production of reproductive brood later in the season. Between two and three weeks after exposure, detectable differences in worker numbers between treated and control colonies began to emerge, persisting until close to the end of the colony cycle (Fig. 1a and Supplementary Table 2d; analysis using a generalized linear mixed-effects model: treatment parameter estimate = −0.28, 95% confidence interval = −0.48 to −0.01; treatment:week interaction parameter estimate = −0.06, 95% confidence interval = −0.11 to −0.01; treatment:week^2 interaction parameter estimate = 0.11, 95% confidence interval = 0.05 to 0.16).

As the colony cycle progressed, negative impacts on the reproductive output of the treated colonies became apparent. Treated and control colonies were equally likely to produce male reproductive offspring, but treated colonies produced significantly fewer males in total (zero-inflated count model, binomial section, treatment parameter estimate = 0.71, 95% confidence interval = −0.67 to 2.09; count section, treatment parameter estimate = −0.54, 95% confidence interval = −0.72 to −0.37; Fig. 2). This difference became apparent from approximately week 9 onwards (Fig. 1b). The dry mass of these males was no different from those produced by control colonies (w_i (null model) = 0.974), indicating that our results cannot be explained by differential investment in reproductive biomass. Neither treated nor control colonies produced an abundance of queens, but control colonies produced more gynes than treated colonies (in total, 36 new gynes from 3 out of 26 control colonies, no new gynes were produced by any of the 25 treated colonies); thus our findings hold when the total number of sexual offspring is analysed (zero-inflated count model, binomial section, treatment parameter estimate = 0.71, 95% confidence interval = −0.67 to 2.09; count section, treatment parameter estimate = −0.64, 95% confidence interval = −0.81 to −0.46). The timing of reproductive onset, queen longevity and colony survival did not differ between control and treated colonies (Extended Data Fig. 2; survival analyses, treatment parameter estimate for reproductive onset = −0.05, 95% confidence interval = −0.41 to 0.31; colony longevity = −0.03, 95% confidence interval = −0.43 to 0.38; queen survival = −0.07, 95% confidence interval = −0.47 to 0.33).

[1]School of Biological Sciences, Royal Holloway University of London, Egham, UK. *e-mail: Harry.Siviter.2016@live.rhul.ac.uk

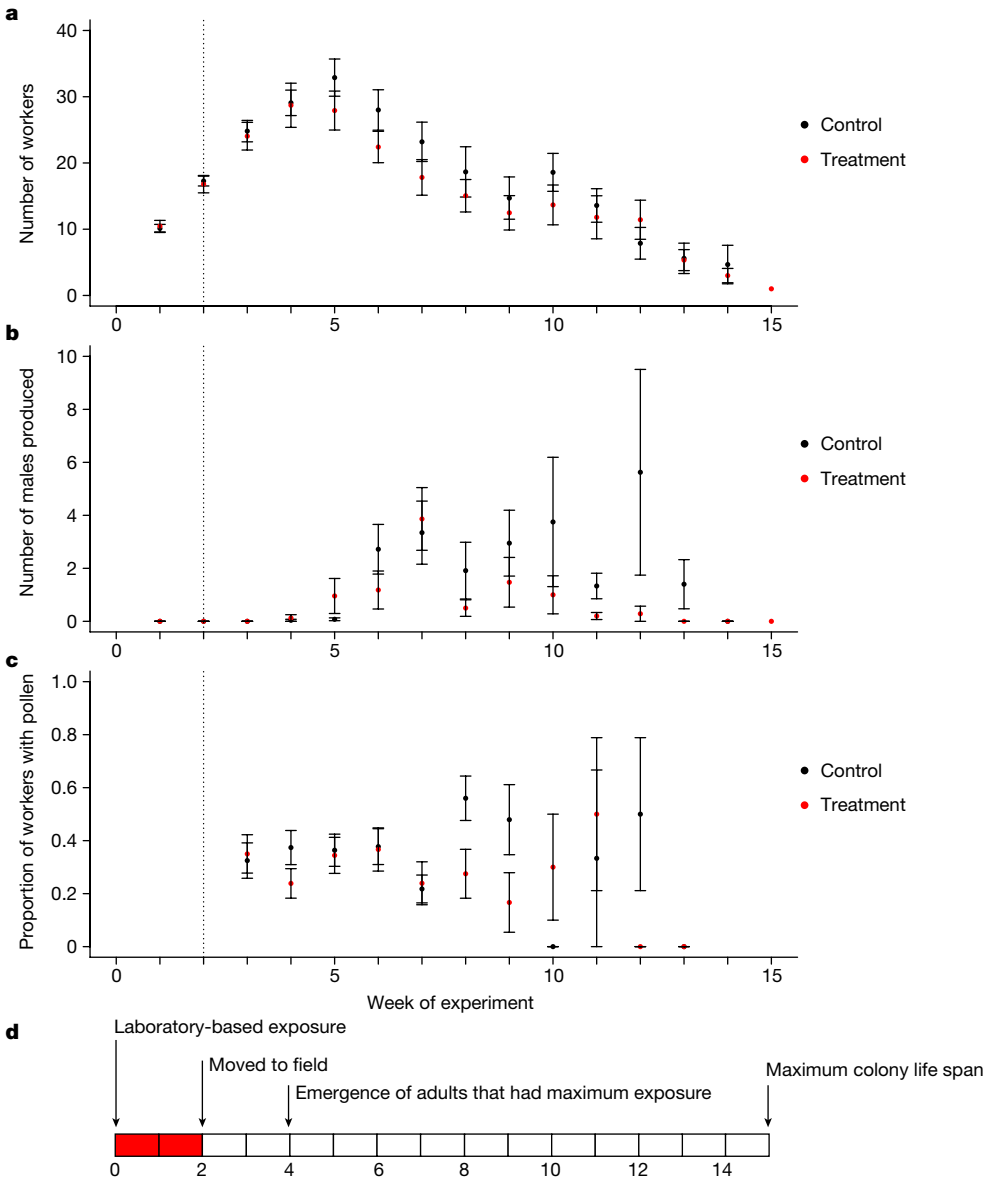


Fig. 1 | The impact of sulfoxaflor exposure on life-history trajectories of bumblebee colonies. a–c, Week-by-week colony field census data. **a**, Number of workers from treated ($n = 26$) and control colonies ($n = 26$). **b**, Number of sexual offspring. **c**, Proportion of workers returning to the colony with pollen for treated and control colonies ($n = 25$ and 26 respectively; reduced sample size for treated colonies reflects the death of one queen in week 2, see Methods). Data are mean ± s.e.m. **d**, Demographic timeline indicates the time points at which the laboratory-based exposure started (the exposure period is indicated in red); the colonies were moved into the field; adults that encountered maximum exposure as larvae should begin to emerge[29] and the maximum lifespan of the colony.

On the basis of the neonicotinoid literature, we considered whether this difference in the production of sexual offspring was mediated through poor provisioning of larvae by foraging workers[9,21], at the time when sexual offspring were developing. However, daytime foraging censuses revealed no significant differences in the relative number of bees returning to control and treated colonies (generalized linear mixed model, treatment parameter estimate = −0.07, 95% confidence interval = −0.32 to 0.19). Similarly, although visual inspection of the data suggested that a lower proportion of workers returned with pollen to pesticide-treated compared to control colonies from week eight onwards (Fig. 1c), this effect did not receive statistical support (generalized linear mixed model, week:treatment interaction parameter estimate = −0.14, 95% confidence interval = −0.29 to 0.001; treatment parameter estimate = 0.46, 95% confidence interval = −0.38 to 1.31) and furthermore occurred too late in the colony cycle to explain the differences in production of male offspring, which became apparent at approximately the same time. We also found no significant differences in the size of pollen loads collected between control and pesticide-treated colonies (Extended Data Fig. 3). Instead, consideration of the timing of differences between control and treated colonies suggests that the effects of sulfoxaflor exposure on reproductive output were mediated by the early drop in worker numbers that began at 2–3 weeks after exposure. Bumblebee worker pupae take approximately 14 days to develop[29], so the onset of deceleration of the growth of the colony workforce corresponds to the eclosion of individuals that had encountered maximum exposure as larvae (Fig. 1d). It remains unclear whether this failure to eclose was driven by direct effects on exposed larvae[30], or indirect effects, perhaps mediated by poor provisioning[9,21] by exposed workers (although note that colonies were provided with pollen and sucrose in the laboratory during this time). In either case, the resultant drop in worker numbers led to differences in the life-history trajectories of control and sulfoxaflor-treated colonies, with consequent effects on the reproductive output of treated colonies[14]. These knock-on effects of early exposure to a small cohort of colony

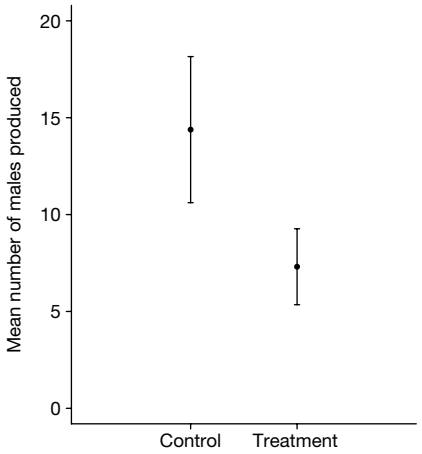


Fig. 2 | Male offspring production. The number of male sexual offspring produced in sulfoxaflor-treated ($n = 25$) and control ($n = 26$) colonies. Data are mean ± s.e.m.

members are entirely consistent with the results of mathematical explorations of stress impacts on bee colonies, which predict that chronic stress at an early stage can push bee colonies beyond a 'tipping point', increasing the likelihood of colony failure[14].

Sulfoxaflor is a systemic pesticide that is soluble in water and is thus transported around plant tissues following foliar or seed application. The likely exposure trajectory of pesticide treatments on crops differs between seed treatments, which deliver prolonged exposure, and spray applications, which deliver a short-term dose that is initially high but typically declines rapidly. Sulfoxaflor, like neonicotinoid-based pesticides, can be administered using both methods, and sulfoxaflor-based products that are used as a seed treatment have recently been developed for crops that attract bees (including oilseed crops)[31]. However, most currently marketed preparations are spray applications. The dosage used in this study is below US Environmental Protection Agency estimates for field-realistic immediate post-spray concentrations of sulfoxaflor in forager-collected nectar, and remains below residual concentrations estimated at 10 days after spray application (the maximum period for which data are available; concentration range over the whole period: 5.41–46.97μg active ingredient (a.i.) per kg, application rate: 0.045 pounds (0.020 kg) of active ingredient per acre applied twice[27]; Extended Data Fig. 1a, b). Note that our treatment protocol is particularly conservative in that our nascent colonies were fed untreated pollen in addition to the syrup provided, potentially producing underestimates of the effects on larvae. Post-spray sulfoxaflor residues in pollen have been documented to be more than tenfold higher than those in forager-collected nectar (Extended Data Fig. 1a, b), ranging from 510.95 to 50.12 μg a.i. per kg over the same post-spray period[27]. Mitigation measures can be used to reduce bee exposure to sulfoxaflor when used as spray treatments (for example, spray application to crops that attract bees during bloom is prohibited by law in the United States)[32]. Globally, however, under current usage, such measures are often either absent[33] or limited to product label recommendations to avoid spraying six days before bloom[34]. No such measures are possible for those products that have been developed as a seed treatment[31].

The impact of sulfoxaflor identified here can be compared with previous experiments that focused on exposure to neonicotinoids. For example, bumblebee colonies placed next to oilseed rape fields that were treated with neonicotinoids showed a 71% reduction in the mean number of queen cocoons found within the nest[6] and a 32–36% reduction in the mean number of males and/or workers produced[7]. Similarly, colonies foraging next to thiacloprid-treated raspberry crops had a 46% reduction in reproductive output[35] and commercial bumblebee colonies exposed to imidacloprid for a period of two weeks had an 85% reduction in the number of new queens produced[8]. Here, we found that sulfoxaflor-exposed colonies had a 54% reduction in the total number of sexual offspring produced compared with control colonies, suggesting that from the perspective of wild pollinators, sulfoxaflor exposure could lead to similar environmental impacts as neonicotinoids if used on crops that attract bees in the absence of evidence-based legislation.

Sulfoximine-based pesticides are a newly emerging class of product, but are already licensed in many countries worldwide, including China[3], Canada[28] and Australia[36]. Within the European Union, where the use of certain neonicotinoids is now banned for open-field crops, substances containing sulfoxaflor as an active ingredient have been assessed by the European Food Safety Authority[37] and approval has been granted for use in five member states, and applications from seven more member states are currently in progress[38]. Our results provide pre-emptive evidence that, if exposure at equivalent dosages to those used in our study occurs via bee-attractive crops before or during bloom, either through spray or seed treatment applications, these products could pose a substantial risk to pollinators. The effects that we identified were the longer-term outcome of initial short-term exposure, and were only detected by monitoring the full colony cycle. Bans and restrictions on neonicotinoid-based pesticides have largely been implemented to protect important pollinators such as bees, following years of widespread use with potential long-term population-level consequences. To avoid a situation in which pesticides such as neonicotinoids are replaced by products that are similarly contentious, regulatory bodies should move towards an evidence-based approach that assesses both the lethal and sub-lethal consequences of novel insecticides such as sulfoxaflor on non-target organisms, and incentivises integrated pest-management approaches before products are licensed for use[39].

Online content

Any Methods, including any statements of data availability and Nature Research reporting summaries, along with any additional references and Source Data files, are available in the online version of the paper at https://doi.org/10.1038/s41586-018-0430-6

Received: 22 February 2018; Accepted: 20 July 2018;
Published online 15 August 2018.

1. Foley, J. A. et al. Global consequences of land use. *Science* **309**, 570–574 (2005).
2. Tilman, D., Cassman, K. G., Matson, P. A., Naylor, R. & Polasky, S. Agricultural sustainability and intensive production practices. *Nature* **418**, 671–677 (2002).
3. Simon-Delso, N. et al. Systemic insecticides (neonicotinoids and fipronil): trends, uses, mode of action and metabolites. *Environ. Sci. Pollut. Res. Int.* **22**, 5–34 (2015).
4. Woodcock, B. A. et al. Impacts of neonicotinoid use on long-term population changes in wild bees in England. *Nat. Commun.* **7**, 12459 (2016).
5. Tsvetkov, N. et al. Chronic exposure to neonicotinoids reduces honey bee health near corn crops. *Science* **356**, 1395–1397 (2017).
6. Rundlöf, M. et al. Seed coating with a neonicotinoid insecticide negatively affects wild bees. *Nature* **521**, 77–80 (2015).
7. Woodcock, B. A. et al. Country-specific effects of neonicotinoid pesticides on honey bees and wild bees. *Science* **356**, 1393–1395 (2017).
8. Whitehorn, P. R., O'Connor, S., Wackers, F. L. & Goulson, D. Neonicotinoid pesticide reduces bumble bee colony growth and queen production. *Science* **336**, 351–352 (2012).
9. Feltham, H., Park, K. & Goulson, D. Field realistic doses of pesticide imidacloprid reduce bumblebee pollen foraging efficiency. *Ecotoxicology* **23**, 317–323 (2014).
10. Pisa, L. et al. An update of the worldwide integrated assessment (WIA) on systemic insecticides. Part 2: impacts on organisms and ecosystems. *Environ. Sci. Pollut. Res. Int.* https://doi.org/10.1007/s11356-017-0341-3 (2017).
11. Brown, M. J. F. et al. A horizon scan of future threats and opportunities for pollinators and pollination. *PeerJ* **4**, e2249 (2016).
12. *Commission Implementing Regulation (EU) 2015/1295 of 27 July 2015 approving the active substance sulfoxaflor, in accordance with Regulation (EC) No 1107/2009 of the European Parliament and of the Council concerning the placing of plant protection products and amending the Annex to Commission Implementing Regulation (EU) No 540/2011* https://publications.europa.eu/en/publication-detail/-/publication/e5aad62c-35b1-11e5-9f85-01aa75ed71a1/language-en (European Commission, 2015).
13. Milner, A. M. & Boyd, I. L. Toward pesticidovigilance. *Science* **357**, 1232–1234 (2017).
14. Bryden, J., Gill, R. J., Mitton, R. A. A., Raine, N. E. & Jansen, V. A. A. Chronic sublethal stress causes bee colony failure. *Ecol. Lett.* **16**, 1463–1469 (2013).
15. Baron, G. L., Jansen, V. A. A., Brown, M. J. F. & Raine, N. E. Pesticide reduces bumblebee colony initiation and increases probability of population extinction. *Nat. Ecol. Evol.* **1**, 1308–1316 (2017).

16. Bass, C., Denholm, I., Williamson, M. S. & Nauen, R. The global status of insect resistance to neonicotinoid insecticides. *Pestic. Biochem. Physiol.* **121**, 78–87 (2015).
17. Zhu, Y. et al. Discovery and characterization of sulfoxaflor, a novel insecticide targeting sap-feeding pests. *J. Agric. Food Chem.* **59**, 2950–2957 (2011).
18. Longhurst, C. et al. Cross-resistance relationships of the sulfoximine insecticide sulfoxaflor with neonicotinoids and other insecticides in the whiteflies *Bemisia tabaci* and *Trialeurodes vaporariorum*. *Pest Manag. Sci.* **69**, 809–813 (2013).
19. Perry, T. et al. Effects of mutations in *Drosophila* nicotinic acetylcholine receptor subunits on sensitivity to insecticides targeting nicotinic acetylcholine receptors. *Pestic. Biochem. Physiol.* **102**, 56–60 (2012).
20. Sparks, T. C. et al. Differential metabolism of sulfoximine and neonicotinoid insecticides by *Drosophila melanogaster* monooxygenase CYP6G1. *Pestic. Biochem. Physiol.* **103**, 159–165 (2012).
21. Gill, R. J., Ramos-Rodriguez, O. & Raine, N. E. Combined pesticide exposure severely affects individual- and colony-level traits in bees. *Nature* **491**, 105–108 (2012).
22. Henry, M. et al. A common pesticide decreases foraging success and survival in honey bees. *Science* **336**, 348–350 (2012).
23. Baron, G. L., Raine, N. E. & Brown, M. J. F. General and species-specific impacts of a neonicotinoid insecticide on the ovary development and feeding of wild bumblebee queens. *Proc. R. Soc. B* **284**, 20170123 (2017).
24. Samuelson, E. E. W., Chen-Wishart, Z. P., Gill, R. J. & Leadbeater, E. Effect of acute pesticide exposure on bee spatial working memory using an analogue of the radial-arm maze. *Sci. Rep.* **6**, 38957 (2016).
25. Siviter, H., Koricheva, J., Brown, M. J. F. & Leadbeater, E. Quantifying the impact of pesticides on learning and memory in bees. *J. Appl. Ecol.* https://doi.org/10.1111/1365-2664.13193 (2018).
26. Stanley, D. A. et al. Neonicotinoid pesticide exposure impairs crop pollination services provided by bumblebees. *Nature* **528**, 548–550 (2015).
27. United States Environmental Protection Agency. 2016 Addendum to the Environmental Fate and Ecological Risk Assessment for Sulfoxaflor Registration (2016).
28. Pest Management Regulatory Agency. *Registration Decision RD2016–12, Sulfoxaflor* https://canada.ca/en/health-canada/services/consumer-product-safety/reports-publications/pesticides-pest-management/decisions-updates/registration-decision/2016/sulfoxaflor-rd2016-12.html (2016).
29. Alford, D. V. *Bumblebees* (Davis-Poynter, London, 1975).
30. Wu, J. Y., Anelli, C. M. & Sheppard, W. S. Sub-lethal effects of pesticide residues in brood comb on worker honey bee (*Apis mellifera*) development and longevity. *PLoS ONE* **6**, e14720 (2011).
31. Syngenta Canada. *Visivio Seed Treatment* https://www.syngenta.ca/Productsdetail/Visivio (2018).
32. Centner, T. J., Brewer, B. & Leal, I. Reducing damages from sulfoxaflor use through mitigation measures to increase the protection of pollinator species. *Land Use Policy* **75**, 70–76 (2018).
33. Dow AgroSciences South Africa. *Closer 240 SC* http://www.dowagro.com/en-za/southafrica/products/insecticide/closer-240-sc (2018).
34. Dow AgroSciences Ireland. *Closer* http://www.dowagro.com/en-ie/ireland/product-finder/insecticide/closer (2018).
35. Ellis, C., Park, K. J., Whitehorn, P., David, A. & Goulson, D. The neonicotinoid insecticide thiacloprid impacts upon bumblebee colony development under field conditions. *Environ. Sci. Technol.* **51**, 1727–1732 (2017).
36. Australian Pesticides and Veterinary Medicines Authority. *Gazette No. 13* https://apvma.gov.au/node/10401 (2013).
37. European Food Safety Authority. Conclusion on the peer review of the pesticide risk assessment of the active substance sulfoxaflor. *EFSA J.* **12**, 3692 (2014).
38. European Commission. *EU Pesticides Database* http://ec.europa.eu/food/plant/pesticides/eu-pesticides-database/public/?event=activesubstance.detail&language=EN&selectedID=2282 (2018).
39. Sanchez-Bayo, F. & Tennekes, H. A. Assessment of ecological risks of agrochemicals requires a new framework. *Environ. Risk Assess. Remediat.* **1**, 1–9 (2017).

Acknowledgements We thank E. Bailes, J. Bagi, J. Blackwell, A. Folly, C. Martin, A. Samuelson, K. Liu, M. Burke, S. Cobacho Jimenez and E. Wrake for technical assistance in the field and laboratory; Natural England and the Crown Estate for permission to collect bumblebees from Windsor Great Park. H.S. was supported by a Royal Holloway University of London Reid PhD Scholarship and by contributions from High Wickham Beekeeper's Association. This project has received funding from the European Horizon 2020 research and innovation programme under grant agreement no.773921. E.L. was supported by European Research Council Starting Grant BeeDanceGap (638873).

Reviewer information *Nature* thanks R. Paxton, N. E. Raine and the other anonymous reviewer(s) for their contribution to the peer review of this work.

Author contributions H.S. conceived the experiment and all authors contributed to the design of the study. H.S. ran the experiments, H.S. and E.L. performed the statistical analysis and all authors contributed to the writing of the manuscript.

Competing interests The authors declare no competing interests.

Additional information
Extended data is available for this paper at https://doi.org/10.1038/s41586-018-0430-6.
Supplementary information is available for this paper at https://doi.org/10.1038/s41586-018-0430-6.
Reprints and permissions information is available at http://www.nature.com/reprints.
Correspondence and requests for materials should be addressed to H.S.
Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

METHODS

Exposure regime. Sulfoxaflor-based preparations have been developed for use on a wide range of bee-attractive crops that flower at varying times of the year. The regime used in our study most closely mimics spring-flowering crops in temperate environments, allowing comparison with similar neonicotinoid-based studies[6,7,15] that also exposed colonies for a short period during the early growth phase.

Preparations containing sulfoxaflor as an active ingredient are currently most commonly applied as a foliar spray. We thus based our pesticide concentrations on the best available information from a realistic and bee-relevant spray experiment reported by the US Environmental Protection Agency (EPA), in which sulfoxaflor was applied to a cotton crop at an application rate of 2×0.045 pounds of active ingredient per acre. Under this application regime, mean sulfoxaflor residue levels in honeybee-collected nectar did not drop below 5 μg a.i. per kg over an 11-day period[27] (the maximum period for which data are available; Extended Data Fig. 1a). We are confident that our exposure is conservative, because (a) in the same experiment, pollen residue levels did not drop below 50 μg a.i. per kg[3,27] (Extended Data Fig. 1b), while we provided all colonies with untreated pollen ad libitum; and (b) this application rate is similar to label recommendations for at least some sulfoxaflor-based products[33]. A second study has also measured residues (in cucumber), but application rates were 1.5 times above recommended usage, and the relevance of this experiment for bees is unclear as the cucumber tissue that was sprayed and sampled was not described[40].

In terms of current usage, our data are most relevant to sulfoxaflor preparations when sprayed on crops immediately before or during bloom (note that this practice has recently been reviewed and prohibited in the United States[27]). Although some product labels recommend avoidance of spraying six days before bloom[34], this ignores experimental data showing that residues could remain present in pollen at levels that we show to have sub-lethal impacts after this six-day period[27] (Extended Data Fig. 1d). Other labels allow spraying during bloom at night[33]. To the best of our knowledge, no data are currently available on field-realistic residues for seed treatment preparations that have been developed for use on oilseed crops and are already available in some markets[28].

Queen rearing. In total, 332 bumblebee (*Bombus terrestris audax*) queens were caught between the 28 February and the 23 March 2017 in Windsor Great Park, Surrey, UK. Chilled queens were transported to the laboratory, where their faeces were microscopically examined for parasites (*Nosema* spp., *Apicystis bombi*, *Sphaerularia bombi* and *Crithidia bombi*; 400× magnification). Parasitized individuals ($n = 54$) were removed from the experiment. A second parasite screening was repeated after one week (29 further queens were removed, $n = 249$ queens remained).

Queens were placed in rearing boxes (67 mm (width) by 127 mm (length) by 50 mm (depth); Allied Plastics) and were provided with a gravity feeder containing an ad libitum supply of 1.8 M sucrose solution (changed weekly; Thorne) and a pollen ball (changed twice weekly, unless the queen was laying eggs in which case more pollen was added; Biobest). Each queen was housed in a dark/red-lit room maintained at 26 °C and 50–60% relative humidity. Queens that did not produce eggs after eight weeks were removed from the experiment ($n = 107$). Once a queen had produced at least six workers, the colony was moved into a wooden nest box (280 mm (width) by 320 mm (length) by 160 mm (depth)) and randomly assigned to a treatment group (see 'Pesticide exposure'). The time taken to reach this stage varied but was on average 7.2 weeks (±s.d. of 1.5 weeks). On transfer, the queens underwent a final parasite screening (2 queens removed). Two queens died before transfer, therefore, 52 colonies reached this stage. The use of colonies from wild-caught queens is a key feature of our experimental design that enabled us to (a) have a complete overview of the lifecycle of these colonies (both in the laboratory and the field, see below), and (b) use colonies with a life history that was adapted to the local environment.

Pesticide exposure. Prior to pesticide exposure, colonies were allocated randomly to control and treatment groups and paired for size according to the number of workers present (mean ± s.d. = 8.43 ± 1.87). Each colony was then provided with an ad libitum supply of either 1.8 M sucrose solution containing 5 μg dm^{-3} (5 ppb) sulfoxaflor (derived from a stock solution of 1 g dm^{-3} in acetone; Greyhound Chromatography and Allied Chemicals) or 1.8 M sucrose containing an equivalent concentration of acetone but no sulfoxaflor for a two-week period. Sucrose solution was weighed on placement in and removal from the colony; no differences in consumption were found between treatment groups (w_i (null model) = 0.985). During the exposure period, we recorded the number of workers produced, colony mass and the number of dead workers on a weekly basis. One queen died during the exposure period, thus 51 colonies were present at the start of the field experiment ($n = 26$ control colonies and $n = 25$ pesticide-treated colonies).

Field placement. After two weeks of exposure in the laboratory, colonies were moved into the field. Nest boxes were placed within plastic field boxes (440 mm (width) by 710 mm (length) by 310 mm (depth); Really Useful Box) containing insulation wrap (Thermawrap) and aluminium foil, and placed at locations around the Royal Holloway University of London campus, Egham, UK (45 ha; Extended Data Fig. 4). Paired colonies were matched for location within the campus, and were positioned at least 20 m from one another to reduce drifting. Each colony entrance was demarcated by a distinctive visual pattern. Colonies were placed in discreet, shaded and southeast-facing locations, and secured with a ratchet strap to avoid badger damage. To prevent usurpation attempts from other queens and social parasite species *(Bombus vestalis)*, queen excluders were placed on each colony. Upon initial placement in the field, the colonies were supplied with a gravity feeder containing 46 g 1.8 M sucrose solution, after which they received no further food supplements. The process of field placement was staggered over six weeks (10 April to 21 May 2017) owing to variation in the date at which queens were initially caught. The week of placement was included as a predictor in each statistical analysis (see 'Statistical analysis').

Data collection. We combined methodological approaches from previous studies on the effects of neonicotinoids on bumblebees[8,21], as well as studies on bumblebee life history[41] to maximize our measurement of both impacts and potential mechanisms. We conducted censuses every night such that each colony was visited once per week, between the hours of 21:30 and 04:00. Using a red-light torch, we recorded the number of live workers (average of three counts), dead workers, males and new queens. We also recorded the state of the original queen (dead or alive), the presence of gyne larvae and/or pupae, the presence of worker larvae and/or pupae, the number of pollen and nectar pots containing stores, and the mass of the colony (average of three recordings; EM-30KAM balance, A&D Instruments). In cases in which the wax covering prevented observation, we peeled it back in order to conduct the count. Weekly censuses continued until moribundity, defined as either a live queen and three or fewer workers, or no queen and 10 workers or fewer[42]. After the experiment, all sexual offspring that had been found in the colonies ($n = 600$) were dried for 72 h and weighed (accuracy of ±0.001 g).

All 51 colonies were also visited during daylight hours twice per week. Colony traffic (number of bees entering and leaving the nest) was recorded during 10-min counts, once between 9:00 and 13:00 and once between 14:00 and 18:00. We also recorded whether returning workers had large (pollen basket was over-flowing) or small (pollen enclosed within pollen basket) pollen loads relative to their body size. Control and pesticide pairs were always observed directly after one another, in a random order. The average daily temperature, humidity and total rainfall were obtained from a local weather station (https://wunderground.com).

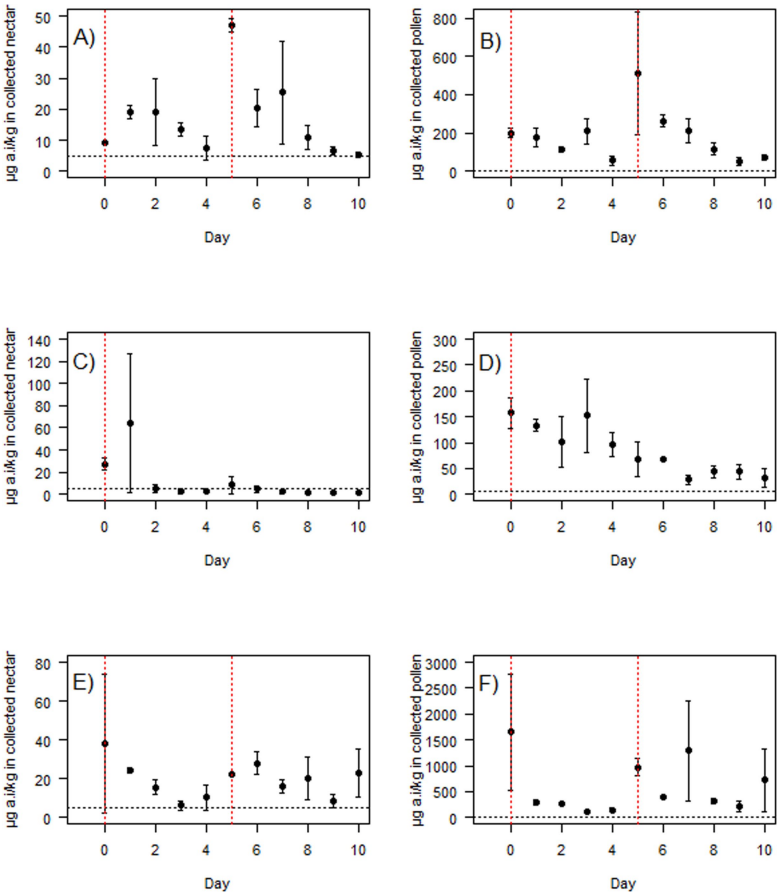
Statistical analyses. We used an information theoretical model selection approach. For each response variable, the initial candidate set included a full model and all subsets, including a null model. Reported parameter estimates and confidence intervals are based on full-set averaging of the 95% confidence set (that is, the set of models with cumulative Akaike weight ≥0.95). Model types, error structuring, a list of parameters included within each model and parameter estimates are provided in Supplementary Tables 1, 2. In brief, to analyse the number of workers produced per week, we used a generalized linear model (glmer; Poisson error structure) with colony nested within the pair as a random factor, and the week of initial field placement (week started), treatment, week of experiment and a two-way interaction between treatment and week of experiment as fixed factors. Because the number of workers increased to a maximum and then decreased for each colony, 'week of experiment' was modelled as a quadratic factor (ΔAIC between full linear and full quadratic model: 1206.40). Many colonies did not produce sexual offspring, so we used zero-inflated generalized linear models (zeroinfl) to analyse the differences in both the overall number of sexual offspring and the number of males produced by colonies, with the week of initial field placement, treatment and their interaction as predictors. The number of workers returning to the nest was analysed using a zero-inflated generalized linear model (glmmadmb; negative binomial error structure) in which treatment, week started, colony week and temperature were included as fixed factors and colony as a random factor. The proportion of workers returning with pollen was also analysed using a generalized linear model (glmmadmb; binomial error structure) with treatment, colony week and their interaction, week started, temperature and time of day included as fixed factors and colony/pair included as a random factor. Week of reproductive onset and queen survival were analysed using a Cox proportional hazards survival analysis that contained treatment and week started as fixed factors. All analyses were conducted in R studio (version 1.0.136) using the R packages pscl[43], lme4[44], glmm[45], MuMin[46] survival[47] and glmmadmb[48].

Reporting summary. Further information on experimental design is available in the Nature Research Reporting Summary linked to this paper.

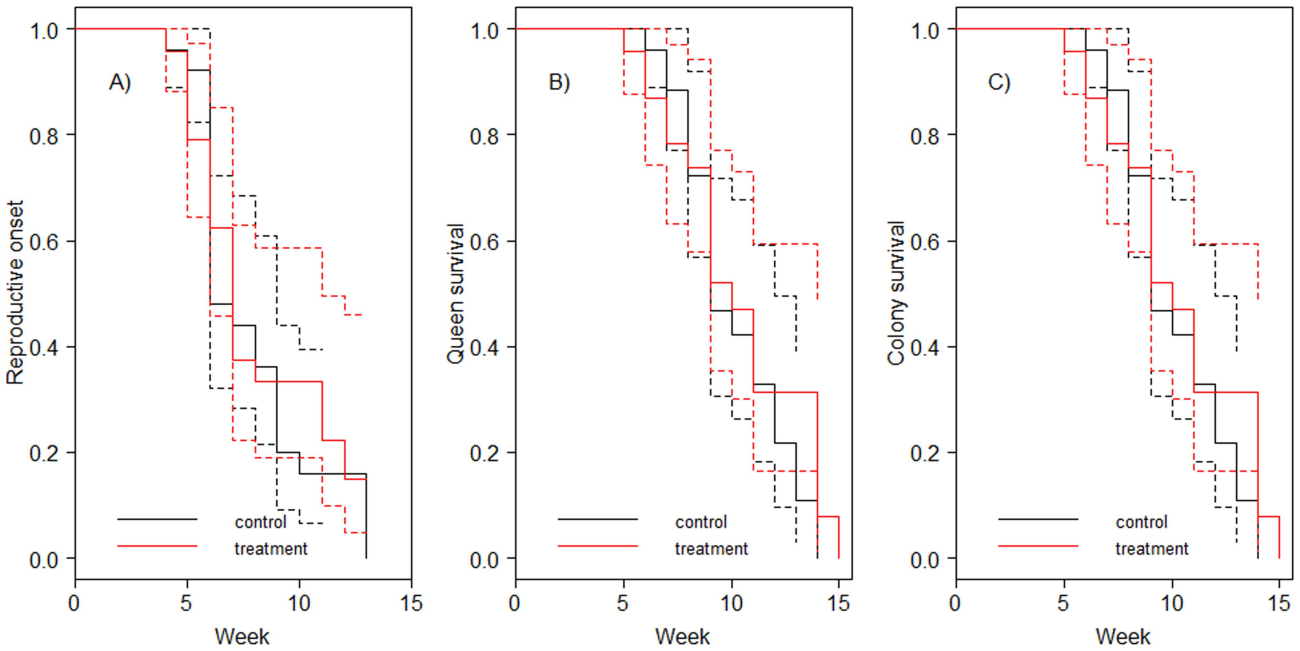
Data availability. The full dataset is available as an open science framework project (https://osf.io/acrsy/).

40. Xu, J. et al. Determination of sulfoxaflor residues in vegetables, fruits and soil using ultra-performance liquid chromatography/tandem mass spectrometry. *Anal. Methods* **4**, 4019–4024 (2012).

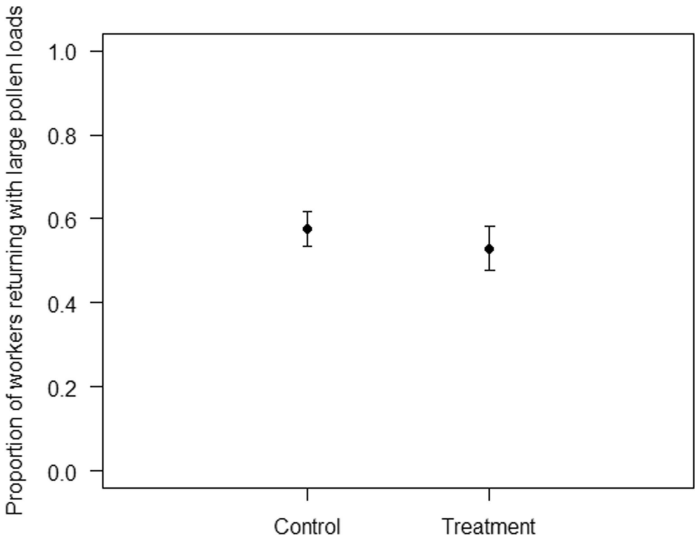
41. Baer, B. & Schmid-Hempel, P. Experimental variation in polyandry affects parasite loads and fitness in a bumble-bee. *Nature* **397**, 151–154 (1999).
42. Samuelson, A. E., Gill, R. J., Brown, M. J. F. & Leadbeater, E. Lower bumblebee colony reproductive success in agricultural compared with urban environments. *Proc. R. Soc. B* **285**, 20180807 (2018).
43. Jackman, S. pscl: Classes and Methods for R Developed in the Political Science Computational Laboratory R package version 1.04.1 https://cran.r-project.org/web/packages/pscl/index.html (2011).
44. Bates, D., Mächler, M., Bolker, B. & Walker, S. Fitting linear mixed-effects models using lme4. *J. Stat. Softw.* **67**, 1–48 (2015).
45. Geyer, C. J. On the convergence of Monte Carlo maximum likelihood calculations. *J. R. Stat. Soc. B* **56**, 261–274 (1994).
46. Burnham, K. P. & Anderson, D. R. *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach 2nd edn* (Springer-Verlag, New York, 2002).
47. Therneau, T. & Grambsch, P. *Modeling Survival Data: Extending the Cox Model* (Springer, New York, 2000).
48. Kristensen, K., Nielsen, A., Berg, C. W., Skaug, H. & Bell, B. M. TMB : automatic differentiation and Laplace approximation. *J. Stat. Softw.* **70**, 1–21 (2016).



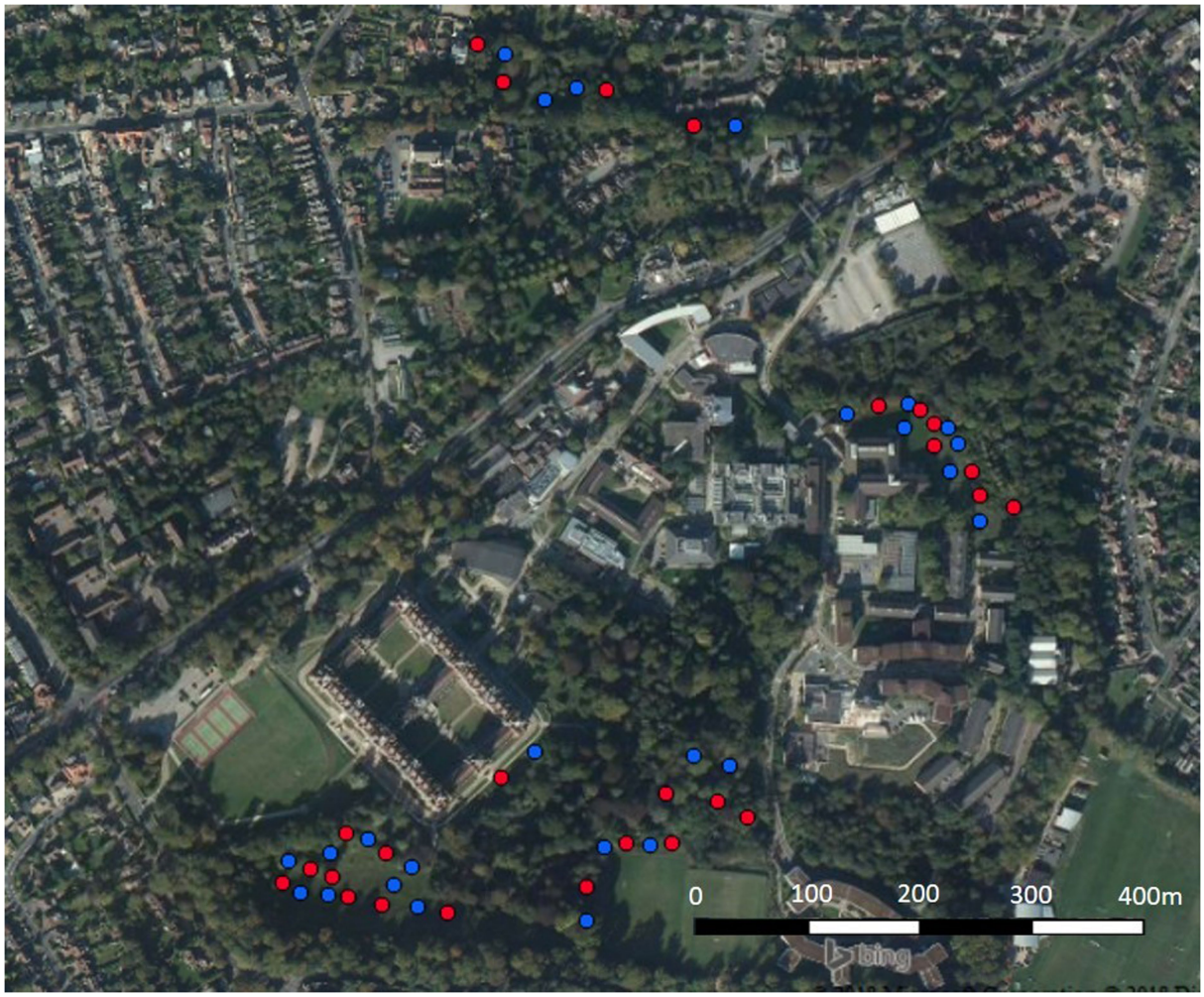
Extended Data Fig. 1 | Concentrations of sulfoxaflor in forager-collected resources from a USA EPA cotton study. Mean µg of active ingredient (a.i.) per kg (mean ± s.e.m.) found in the nectar (**a**, **c**, **e**) and pollen (**b**, **d**, **f**) of honeybees foraging on cotton crops sprayed with sulfoxaflor. Note the differences in *y*-axis scale between graphs, owing to considerably higher concentrations in pollen. Red lines indicate spray application. Dosage: twice over ten days at 0.045 pounds a.i. per acre (**a**, **b**); once over ten days at 0.045 pounds a.i. per acre (**c**, **d**); twice over ten days at 0.089 pounds a.i. per acre (**e**, **f**). The black dotted horizontal line indicates the equivalent amount of sulfoxaflor (5 ppb) that was fed to sulfoxaflor-treated colonies in sucrose in our experiment. Data are means from two hives; number of individual bees sampled is not published[27].



Extended Data Fig. 2 | Timing of colony life-history events. a–c, The probability of reproductive onset (**a**), queen survival (**b**) and colony survival (**c**) for control ($n = 26$) and sulfoxaflor-treated ($n = 25$) colonies (± confidence intervals).



Extended Data Fig. 3 | Pollen foraging. The proportion (mean ± s.e.m.) of foragers returning to the nest with large pollen loads, for control ($n = 25$) and pesticide-treated ($n = 22$) colonies (note that not all of the colonies in the experiment had pollen foragers).



Extended Data Fig. 4 | Distribution of colonies across the Royal Holloway Campus. Blue dots indicate control colonies; red dots indicate treated colonies. Grid reference: TQ000706; Imagery © Google, Map Data © 2018 Google.

LETTER

https://doi.org/10.1038/s41586-018-0455-x

The genome of the offspring of a Neanderthal mother and a Denisovan father

Viviane Slon[1,7]*, Fabrizio Mafessoni[1,7], Benjamin Vernot[1,7], Cesare de Filippo[1], Steffi Grote[1], Bence Viola[2,3], Mateja Hajdinjak[1], Stéphane Peyrégne[1], Sarah Nagel[1], Samantha Brown[4], Katerina Douka[4,5], Tom Higham[5], Maxim B. Kozlikin[3], Michael V. Shunkov[3,6], Anatoly P. Derevianko[3], Janet Kelso[1], Matthias Meyer[1], Kay Prüfer[1] & Svante Pääbo[1]*

Neanderthals and Denisovans are extinct groups of hominins that separated from each other more than 390,000 years ago[1,2]. Here we present the genome of 'Denisova 11', a bone fragment from Denisova Cave (Russia)[3] and show that it comes from an individual who had a Neanderthal mother and a Denisovan father. The father, whose genome bears traces of Neanderthal ancestry, came from a population related to a later Denisovan found in the cave[4–6]. The mother came from a population more closely related to Neanderthals who lived later in Europe[2,7] than to an earlier Neanderthal found in Denisova Cave[8], suggesting that migrations of Neanderthals between eastern and western Eurasia occurred sometime after 120,000 years ago. The finding of a first-generation Neanderthal–Denisovan offspring among the small number of archaic specimens sequenced to date suggests that mixing between Late Pleistocene hominin groups was common when they met.

Neanderthals and Denisovans inhabited Eurasia until they were replaced by modern humans around 40,000 years ago (40 ka)[9]. Neanderthal remains have been found in western Eurasia[10], whereas physical remains of Denisovans have thus far been found only in Denisova Cave[4–6,11,12], where Neanderthal remains have also been recovered[8]. Although little is known about the morphology of Denisovans, their molars lack the derived traits that are typical of Neanderthals[5,11].

DNA recovered from individuals of both groups suggests that they diverged from each other more than 390 ka[1,2]. The presence of small amounts of Neanderthal DNA in the genome of 'Denisova 3', the first Denisovan individual to be identified[4–6], indicates that the two groups mixed with each other at least once[8]. It has also been shown that Neanderthals mixed with the ancestors of present-day non-Africans around 60 ka[2,8,13], and possibly with earlier ancestors of modern humans[1,14,15]; and that Denisovans mixed with the ancestors of present-day Oceanians and Asians[5,16,17]. Denisovans may furthermore have received gene flow from an archaic hominin that diverged more than a million years ago from the ancestors of modern humans[8].

A fragment of a long bone, 'Denisova 11' (Fig. 1), was identified among over 2,000 undiagnostic bone fragments excavated in Denisova Cave as being of hominin origin using collagen peptide mass fingerprinting[3]. Its mitochondrial (mt)DNA was found to be of the Neanderthal type and direct radiocarbon dating showed it to be more than 50,000 years old[3]. From its cortical thickness, we infer that Denisova 11 was at least 13 years old at death (Extended Data Fig. 1 and Supplementary Information 1). We performed six DNA extractions[18,19] from bone powder collected from the specimen, produced ten DNA libraries[20] from the extracts (Extended Data Table 1 and Supplementary Information 2, 3) and sequenced the Denisova 11 genome to an average coverage of 2.6-fold. The coverage of the X chromosome was similar to that of the autosomes, indicating that Denisova 11 was a female. Using three different methods, we estimate that contaminating present-day human DNA fragments constitute at most 1.7% of the data (Supplementary Information 2).

To determine from which hominin group Denisova 11 originated, we compared the proportions of DNA fragments that match derived alleles from a Neanderthal genome ('Altai Neanderthal', also known as 'Denisova 5') or a Denisovan genome (Denisova 3), both determined from bones discovered in Denisova Cave[6,8], as well as from a present-day African genome (Mbuti)[6] (Supplementary Information 4). At informative sites[1], 38.6% of fragments from Denisova 11 carried alleles matching the Neanderthal genome and 42.3% carried alleles matching the Denisovan genome (Fig. 2a), suggesting that both archaic groups contributed to the ancestry of Denisova 11 to approximately equal extents (Supplementary Information 4). Approximately equal proportions of Neanderthal-like and Denisovan-like alleles are found in each of the ten DNA libraries originating from Denisova 11 but not in libraries from other projects that were prepared, sequenced and processed in parallel, which excludes an accidental mixing of DNA in the laboratory or a systematic error in data processing (Supplementary Information 3).

To estimate the heterozygosity of Denisova 11, we restrict the analysis to transversion polymorphisms to prevent deamination-derived substitutions from inflating the estimates, and find 3.7 transversions per 10,000 autosomal base pairs. This is over four times higher than the heterozygosity of the two Neanderthal (Altai Neanderthal and 'Vindija 33.19') and one Denisovan (Denisova 3) genomes sequenced to date, and similar to the heterozygosity seen in present-day Africans. In fact, the heterozygosity of Denisova 11 is similar to what would be expected if this individual carried one set of chromosomes of Neanderthal origin and one of Denisovan origin, as estimated from the number of differences between randomly sampled DNA fragments from either the Vindija 33.19 or the Altai Neanderthal genome and the Denisova 3 genome (Fig. 2b and Supplementary Information 5).

Denisova 11 could have had approximately equal amounts of Neanderthal and Denisovan ancestry because she belonged to a population with mixed Neanderthal and Denisovan ancestry, or because her parents were each from one of these two groups. To determine which of these two scenarios fits the data best, we considered sites at which the genomes of the Altai Neanderthal and Denisova 3 carry a transversion difference in a homozygous form. At each of these sites, we recorded the alleles carried by two randomly drawn DNA fragments from Denisova 11. Note that in 50% of cases, both fragments will come from the same chromosome, making 50% of heterozygous sites appear homozygous. As a consequence, the expected proportion of apparent heterozygous sites is 50% for a first-generation (F_1) offspring, whereas it is 25% in a population at Hardy–Weinberg equilibrium with mixed ancestry in equal proportions (Supplementary Information 6). We find that in 43.5% of cases, one fragment from Denisova 11 matches the Neanderthal genome and the other matches the Denisovan genome, whereas in 27.3% and 29.2% of cases both fragments match the state

[1]Department of Evolutionary Genetics, Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany. [2]Department of Anthropology, University of Toronto, Toronto, Ontario, Canada. [3]Institute of Archaeology and Ethnography, Russian Academy of Sciences, Novosibirsk, Russia. [4]Max Planck Institute for the Science of Human History, Jena, Germany. [5]Oxford Radiocarbon Accelerator Unit, RLAHA, University of Oxford, Oxford, UK. [6]Novosibirsk State University, Novosibirsk, Russia. [7]These authors contributed equally: Viviane Slon, Fabrizio Mafessoni, Benjamin Vernot. *e-mail: viviane_slon@eva.mpg.de; paabo@eva.mpg.de

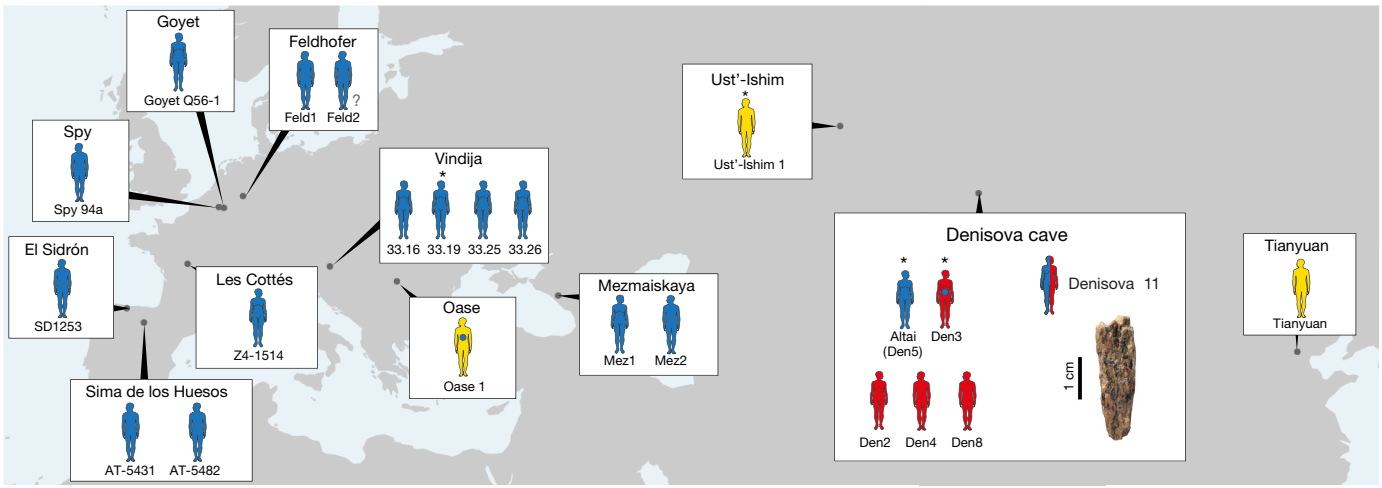


Fig. 1 | Location of Neanderthals, Denisovans and ancient modern humans dated to approximately 40 ka or earlier. Only individuals from whom sufficient nuclear DNA fragments have been recovered to enable their attribution to a hominin group are shown. Full or abbreviated names of specimens are shown near each individual. Blue, Neanderthals; red, Denisovans; yellow, ancient modern humans. Asterisks indicate that the genome was sequenced to high coverage; individuals with an unknown sex are marked with a question mark. Note that Oase 1 has recent Neanderthal ancestry (blue dot) that is higher than the amount seen in non-Africans. Denisova 3 has also been found to carry a small percentage of Neanderthal ancestry. Data were obtained from previous publications[1,2,5–8,11–13,21–24].

seen in the Neanderthal or the Denisovan genome, respectively (Fig. 2c). For comparison, when a low-coverage Neanderthal genome ('Goyet Q56-1')[7] is analysed in the same way, the two fragments match different states in 2.1% of cases, while they both match the Neanderthal state in 90.3% of cases and the Denisovan state in 7.5% of cases (Fig. 2c).

Obviously, the Altai Neanderthal and Denisova 3 are unlikely to be identical to the genomes of the individuals that contributed ancestry to Denisova 11. To take this into account, we used coalescent simulations to estimate the expected proportions of DNA fragments matching a Neanderthal or a Denisovan genome in populations with demographic histories similar to those of the Altai Neanderthal and Denisova 3 (Supplementary Information 6). The proportion of cases in which one of the two DNA fragments sampled from Denisova 11 matches the Neanderthal state and the other the Denisovan state fits the expectation for an F_1 Neanderthal–Denisovan offspring, but not an offspring of two F_1 individuals, an offspring of an F_1 parent and a Neanderthal or a Denisovan parent, nor an individual from a population of mixed ancestry at Hardy–Weinberg equilibrium (Extended Data Fig. 2 and Supplementary Information 6). We conclude that Denisova 11 did not originate from a population carrying equal proportions of Neanderthal and Denisovan ancestry. Rather, she was the offspring of a Neanderthal mother, who contributed her mtDNA, and a Denisovan father.

We next plotted the distribution of sites across the genome, for which Denisova 11 carries an allele matching the Altai Neanderthal genome and a different allele matching the Denisova 3 genome. Such sites are distributed largely uniformly (Fig. 3), as would be expected for an F_1 offspring of Neanderthal and Denisovan parents. To explore the ancestry of the parents of Denisova 11, we looked for regions in the genome that deviate from a pattern consistent with Denisova 11 being an F_1 offspring (Extended Data Fig. 3). Using four tests for enrichment of Denisovan or Neanderthal ancestry, we identify at least five approximately 1-Mb long (0.72–0.95 Mb) regions, all of

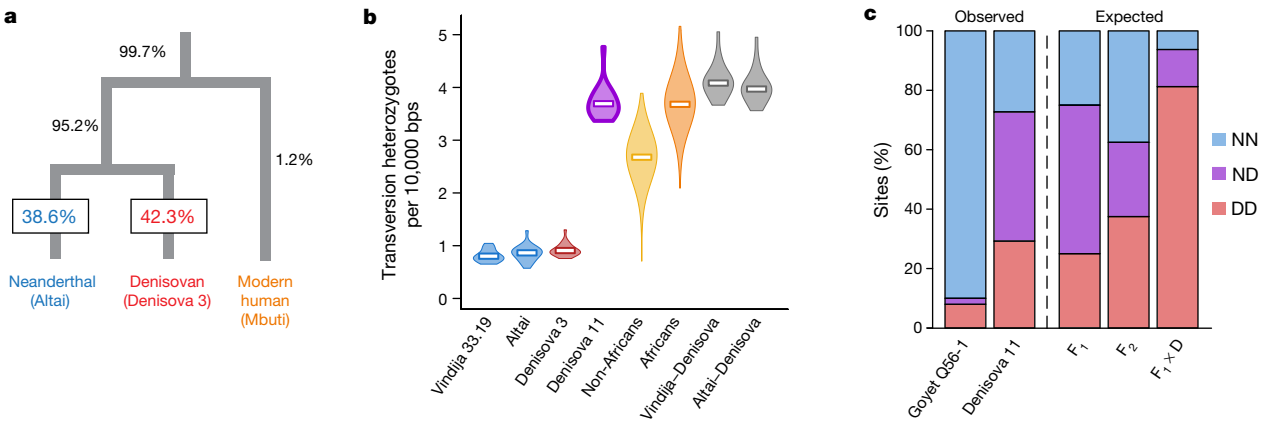


Fig. 2 | Denisova 11 has both Neanderthal and Denisovan ancestry. **a**, Percentage of DNA fragments from Denisova 11 matching derived alleles found on each branch of a tree relating a Neanderthal, a Denisovan and a present-day human genome. **b**, Distribution of heterozygosity per chromosome in two Neanderthals (blue), a Denisovan (red), Denisova 11 (purple) and present-day humans ($n = 235$ non-African individuals (yellow) and $n = 44$ African individuals (orange) from a previous publication[28]), and the expectation for a Neanderthal–Denisovan F_1 offspring (grey). The violins represent the distribution from the minimum and maximum heterozygosity values for the autosomes of each archaic hominin and of present-day humans ($n = 5{,}170$ pairs of chromosomes for non-Africans and $n = 968$ for Africans). White squares represent autosome-wide estimates for the archaic hominins, and the average of estimates across individuals for present-day humans. **c**, Percentage of sites at which two sampled DNA fragments both carry Neanderthal-like alleles (NN, blue), Denisovan-like alleles (DD, red), or one allele of each type (ND, purple); and the expectations for an offspring of a Neanderthal and a Denisovan (F_1), of two F_1 parents (F_2), and of an F_1 and a Denisovan ($F_1 \times D$). The expected proportions for simulated Neanderthal and Denisovan genomes are shown in Extended Data Fig. 2.

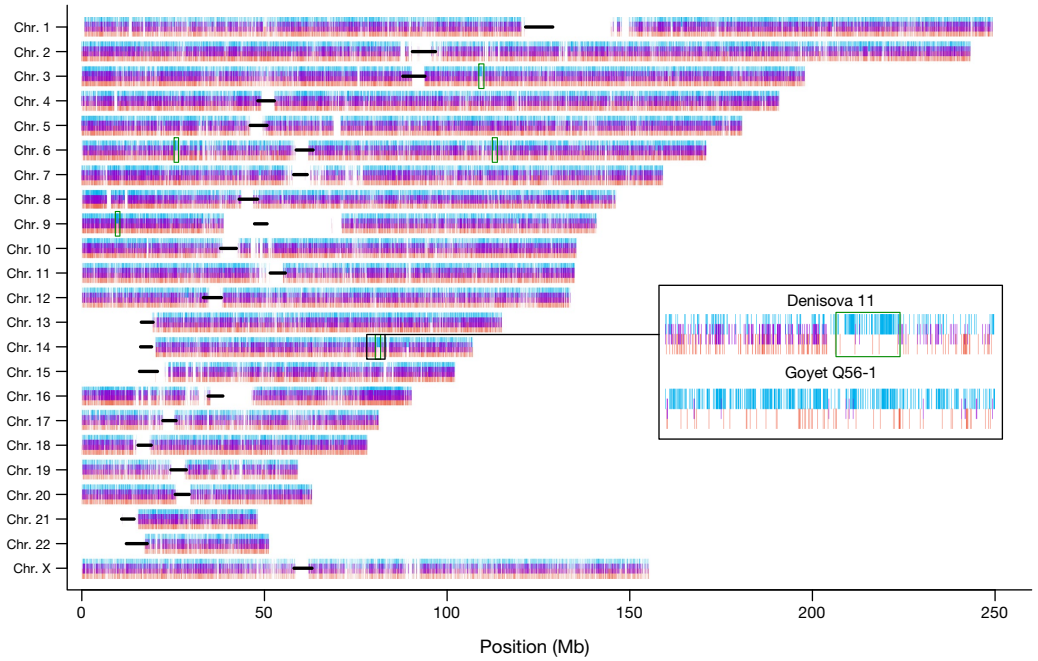


Fig. 3 | Distribution of Neanderthal-like and Denisovan-like alleles across the Denisova 11 genome. Positions for which one randomly drawn DNA fragment matches the Neanderthal genome and another matches the Denisovan genome are marked in purple. Positions are marked in blue if both DNA fragments match the Neanderthal genome and in red if both match the Denisovan genome. Black lines indicate centromeres. The inset shows one region out of five (green boxes) for which both chromosomes carry predominantly Neanderthal-like alleles. For comparison, the distribution of alleles in this region is shown for a Neanderthal genome (Goyet Q56-1).

which are homozygous for Neanderthal ancestry. This suggests that the Denisovan father of Denisova 11 had some Neanderthal ancestry. Given conservative estimates of the size and number of these regions, it is likely that there was more than one Neanderthal ancestor in his genealogy, possibly as far back as 300–600 generations before his lifetime (Supplementary Information 7). Notably, the heterozygosity in the regions of Neanderthal ancestry in Denisova 11 is higher than in the same regions in the genomes of Vindija 33.19 or the Altai Neanderthal, suggesting that the Neanderthals that contributed to the ancestry of Denisova 11's father were from a different population than her mother (Supplementary Information 5).

To explore how the mother of Denisova 11 was related to the two Neanderthals that have been sequenced to high coverage to date, we evaluated the proportions of fragments from Denisova 11 that match derived alleles from either of these two Neanderthal genomes. Denisova 11 shares derived alleles seen in the Altai Neanderthal genome in 12.4% of cases and those present in the Vindija 33.19 genome in 19.6% of cases, showing that the Neanderthal mother of Denisova 11 came from a population that was more closely related to Vindija 33.19 than to the Altai Neanderthal (Supplementary Information 8). We estimate the population split times of Denisova 11's Neanderthal mother from the ancestors of the Altai Neanderthal to approximately 20,000 years (20 kyr) before the time when the Altai Neanderthal lived, and her split time from the ancestors of Vindija 33.19 to around 40 kyr before Vindija 33.19. The population split between the Denisovan father of Denisova 11 and Denisova 3 is estimated to approximately 7 kyr before the latter individual (Supplementary Information 8). In Fig. 4, we present a population scenario that is compatible with these observations as well as with the population split times and molecular estimates of the ages of the three high-coverage archaic genomes[2]. We caution that the age estimates are associated with uncertainties, for example, regarding demography, mutation rates and generation times, and note that additional gene flow events are likely to have affected the population split times. Nevertheless, that a Neanderthal in Siberia who lived approximately 90 ka shared more alleles with Neanderthals who lived at least 20 kyr later in Europe[2,7] than with an earlier Neanderthal from the same cave[8] suggests that eastern Neanderthals spread into Western Europe

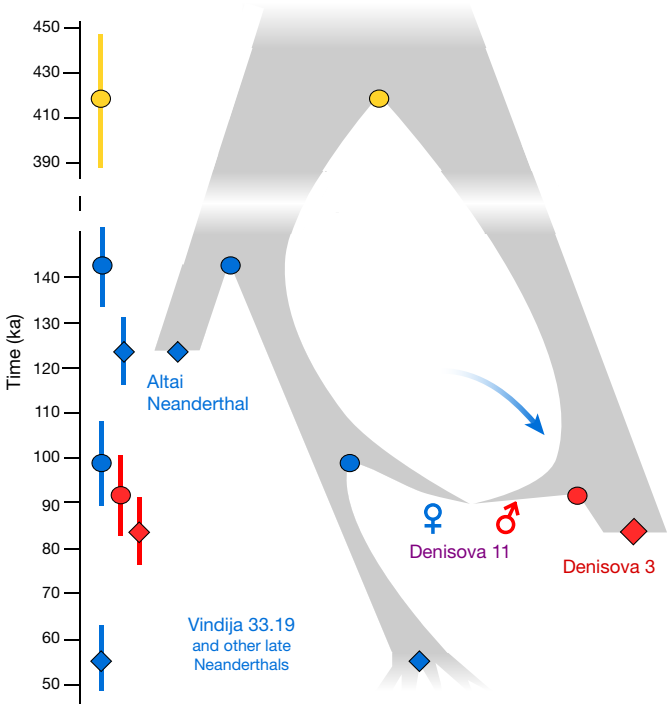


Fig. 4 | Relationships and gene flow events between Neanderthal and Denisovan populations inferred from genome sequences. Diamonds indicate ages of specimens estimated via branch shortening[2]; circles indicate population split times estimated from allele sharing between Denisova 11 and the high-coverage Neanderthal and Denisovan genomes (blue and red) and among the three high-coverage genomes (yellow, from a previous publication[2]). Markers indicate the means of these estimates, error bars indicate 95% confidence intervals based on block jackknife resampling across the genome ($n = 523$ blocks). Note that the confidence intervals do not take the uncertainty with respect to population size, mutation rates or generation times into account. Ages before present are based on a human–chimpanzee divergence of 13 million years[22,29]. The arrow indicates Neanderthal gene flow into Denisovans.

sometime after 90 ka or that western Neanderthals spread to Siberia before that time and partially replaced the local population. These two non-mutually exclusive hypotheses could be tested by sequencing the genomes of early Neanderthals from Western Europe.

In conclusion, the genome of Denisova 11 provides direct evidence for genetic mixture between Neanderthals and Denisovans on at least two occasions: once between her Neanderthal mother and her Denisovan father, and at least once in the ancestry of her Denisovan father. Therefore, of the six individuals from Denisova Cave from whom nuclear DNA is available[5,6,8,11,12], two (Denisova 3 and Denisova 11) show evidence of gene flow between Neanderthals and Denisovans. We note that of the three genomes[21–24] retrieved from modern humans who lived at a time when Neanderthals were present in Eurasia (that is, approximately 40 ka or earlier)[9], one individual—'Oase 1'—had a Neanderthal ancestor four to six generations back in his family tree[23].

It is notable that one direct offspring of a Neanderthal and a Denisovan (Denisova 11) and one modern human with a close Neanderthal relative (Oase 1) have been identified among the few individuals from whom DNA has been retrieved and who lived at the time of overlap of these groups (Fig. 1). In conjunction with the presence of Neanderthal and Denisovan DNA in ancient and present-day people[2,5,8,13,16,17,25–27], this suggests that mixing among archaic and modern hominin groups may have been frequent when they met. However, Neanderthals inhabited western Eurasia[10] whereas Denisovans inhabited yet unknown parts of eastern Eurasia[5,17]. Thus, their zones of overlap may have been restricted in space and time. This, as well as possibly reduced fitness of individuals of mixed ancestry, may explain why Neanderthals and Denisovans remained genetically distinct. By contrast, the spread of modern humans across Eurasia after around 60,000 years ago may have allowed repeated interactions with archaic groups over a wider spatial range. Admixture between them may have resulted in archaic populations becoming partly absorbed into what were probably larger modern human populations[6,8].

Online content

Any methods, additional references, Nature Research reporting summaries, source data, statements of data availability and associated accession codes are available at https://doi.org/10.1038/s41586-018-0455-x.

Received: 20 February; Accepted: 18 June 2018;
Published online 22 August 2018.

1. Meyer, M. et al. Nuclear DNA sequences from the Middle Pleistocene Sima de los Huesos hominins. *Nature* **531**, 504–507 (2016).
2. Prüfer, K. et al. A high-coverage Neandertal genome from Vindija Cave in Croatia. *Science* **358**, 655–658 (2017).
3. Brown, S. et al. Identification of a new hominin bone from Denisova Cave, Siberia using collagen fingerprinting and mitochondrial DNA analysis. *Sci. Rep.* **6**, 23559 (2016).
4. Krause, J. et al. The complete mitochondrial DNA genome of an unknown hominin from southern Siberia. *Nature* **464**, 894–897 (2010).
5. Reich, D. et al. Genetic history of an archaic hominin group from Denisova Cave in Siberia. *Nature* **468**, 1053–1060 (2010).
6. Meyer, M. et al. A high-coverage genome sequence from an archaic Denisovan individual. *Science* **338**, 222–226 (2012).
7. Hajdinjak, M. et al. Reconstructing the genetic history of late Neanderthals. *Nature* **555**, 652–656 (2018).
8. Prüfer, K. et al. The complete genome sequence of a Neanderthal from the Altai Mountains. *Nature* **505**, 43–49 (2014).
9. Higham, T. et al. The timing and spatiotemporal patterning of Neanderthal disappearance. *Nature* **512**, 306–309 (2014).
10. Vandermeersch, B. & Garralda, M. D. in *Continuity and Discontinuity in the Peopling of Europe: One Hundred Fifty Years of Neanderthal Study* (eds Condemi, S. & Weniger, G.-C.) 113–125 (Springer, Dordrecht, 2011).
11. Sawyer, S. et al. Nuclear and mitochondrial DNA sequences from two Denisovan individuals. *Proc. Natl Acad. Sci. USA* **112**, 15696–15700 (2015).
12. Slon, V. et al. A fourth Denisovan individual. *Sci. Adv.* **3**, e1700186 (2017).
13. Green, R. E. et al. A draft sequence of the Neandertal genome. *Science* **328**, 710–722 (2010).
14. Kuhlwilm, M. et al. Ancient gene flow from early modern humans into Eastern Neanderthals. *Nature* **530**, 429–433 (2016).
15. Posth, C. et al. Deeply divergent archaic mitochondrial genome provides lower time boundary for African gene flow into Neanderthals. *Nat. Commun.* **8**, 16046 (2017).
16. Skoglund, P. & Jakobsson, M. Archaic human ancestry in East Asia. *Proc. Natl Acad. Sci. USA* **108**, 18301–18306 (2011).
17. Reich, D. et al. Denisova admixture and the first modern human dispersals into Southeast Asia and Oceania. *Am. J. Hum. Genet.* **89**, 516–528 (2011).
18. Dabney, J. et al. Complete mitochondrial genome sequence of a Middle Pleistocene cave bear reconstructed from ultrashort DNA fragments. *Proc. Natl Acad. Sci. USA* **110**, 15758–15763 (2013).
19. Korlević, P. et al. Reducing microbial and human contamination in DNA extractions from ancient bones and teeth. *Biotechniques* **59**, 87–93 (2015).
20. Gansauge, M. T. & Meyer, M. Single-stranded DNA library preparation for the sequencing of ancient or damaged DNA. *Nat. Protoc.* **8**, 737–748 (2013).
21. Fu, Q. et al. DNA analysis of an early modern human from Tianyuan Cave, China. *Proc. Natl Acad. Sci. USA* **110**, 2223–2227 (2013).
22. Fu, Q. et al. Genome sequence of a 45,000-year-old modern human from western Siberia. *Nature* **514**, 445–449 (2014).
23. Fu, Q. et al. An early modern human from Romania with a recent Neanderthal ancestor. *Nature* **524**, 216–219 (2015).
24. Yang, M. A. et al. 40,000-year-old individual from Asia provides insight into early population structure in Eurasia. *Curr. Biol.* **27**, 3202–3208 (2017).
25. Seguin-Orlando, A. et al. Genomic structure in Europeans dating back at least 36,200 years. *Science* **346**, 1113–1118 (2014).
26. Fu, Q. et al. The genetic history of Ice Age Europe. *Nature* **534**, 200–205 (2016).
27. Sikora, M. et al. Ancient genomes show social and reproductive behavior of early Upper Paleolithic foragers. *Science* **358**, 659–662 (2017).
28. Mallick, S. et al. The Simons Genome Diversity Project: 300 genomes from 142 diverse populations. *Nature* **538**, 201–206 (2016).
29. Scally, A. & Durbin, R. Revising the human mutation rate: implications for understanding human evolution. *Nat. Rev. Genet.* **13**, 745–753 (2012).

Acknowledgements We thank B. Schellbach and A. Weihmann for DNA sequencing; G. Renaud and U. Stenzel for data processing; F. Brock for the computed tomography scans; R. Barr, P. Korlević and C. Zickert for graphics; and M. Slatkin and L. Vigilant for comments on the manuscript. This work was funded by the Max Planck Society; the Max Planck Foundation (grant 31-12LMP Pääbo to S.Pä.); the European Research Council (grant agreements 694707 to S.Pä., 324139 (PalaeoChron) to T.H. and 715069 (FINDER) to K.D.); and the Russian Science Foundation (project no. 14-50-00036 to M.B.K., M.V.S. and A.P.D.).

Reviewer information *Nature* thanks D. Lambert, R. Nielsen and C. Stringer for their contribution to the peer review of this work.

Author contributions V.S. and S.N. performed the laboratory work; V.S., F.M., B.Ve., C.d.F., S.G., M.H., S.Pe., J.K., M.M., K.P. and S.Pä. analysed the genetic data; B.Vi. carried out the morphological analysis; S.B., K.D., T.H., M.B.K., M.V.S. and A.P.D. discovered Denisova 11 and provided archaeological data; V.S., K.P. and S.Pä. wrote the manuscript with input from all authors.

Competing interests The authors declare no competing interests.

Additional information
Extended data is available for this paper at https://doi.org/10.1038/s41586-018-0455-x.
Supplementary information is available for this paper at https://doi.org/10.1038/s41586-018-0455-x.
Reprints and permissions information is available at http://www.nature.com/reprints.
Correspondence and requests for materials should be addressed to V.S. or S.Pä.
Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

METHODS

Sampling and pre-treatment of bone powder. An overview of the laboratory experiments is shown in Extended Data Table 1. Bone powder was removed from the specimen using disposable sterile dentistry drills after the removal of a thin layer of surface material. Six samples were collected, each consisting of approximately 30 mg of bone powder. Because a previous analysis of the bone revealed that it is contaminated with present-day human DNA[3], each sample of bone powder was incubated with 1 ml 0.5% sodium hypochlorite solution as previously described[19] and as indicated in Extended Data Table 1, to reduce the amounts of present-day human and microbial DNA[7,19]. Residual sodium hypochlorite was removed by three consecutive 3-min washes with 1 ml water[19]. One extraction negative control (no powder) was included in each set of extractions.

DNA extraction and DNA library preparation. DNA was extracted using silica columns[18] as previously described[19], and eluted in 50 μl 10 mM Tris-HCl, 1 mM EDTA, 0.05% Tween-20, pH 8.0. Subsequently, 10 μl of each DNA extract (including the extraction negative controls) were used to prepare single-stranded DNA libraries as previously described[19,20]. A library preparation negative control was included in every experiment. Two additional 5-μl aliquots from extracts E3652 and E3655 were used to generate additional libraries (library preparation setup C in Extended Data Table 1), resulting in a total of 10 DNA libraries. The number of DNA molecules in the libraries was estimated by digital droplet PCR[30] or quantitative PCR[20]. Each library was amplified to the plateau while incorporating a pair of unique indexes[31] using 1 μM primers[19,31] and AccuPrime Pfx DNA polymerase (Life Technologies)[32]. Amplification products were purified using the MinElute PCR purification kit (Qiagen) or SPRI technology[33] on a Bravo NGS workstation (Agilent Technologies) as previously described[34]. Indexed DNA libraries were pooled with libraries from other projects. Heteroduplices, which confound DNA separation and concentration measurements in chromatography, were removed from the pools by single cycle amplification using Herculase II Fusion DNA polymerase (Agilent Technologies)[32] with primers IS5 and IS6[35]. Prior to deeper sequencing of libraries R5507, R5509, R9880, R9881, R9882, R9883 and R9873, heteroduplices were removed from each library separately. The concentration of DNA in each pool or each individual library, respectively, was determined using the electrophoresis system implemented on the DNA-1000 chip (Agilent Technologies).

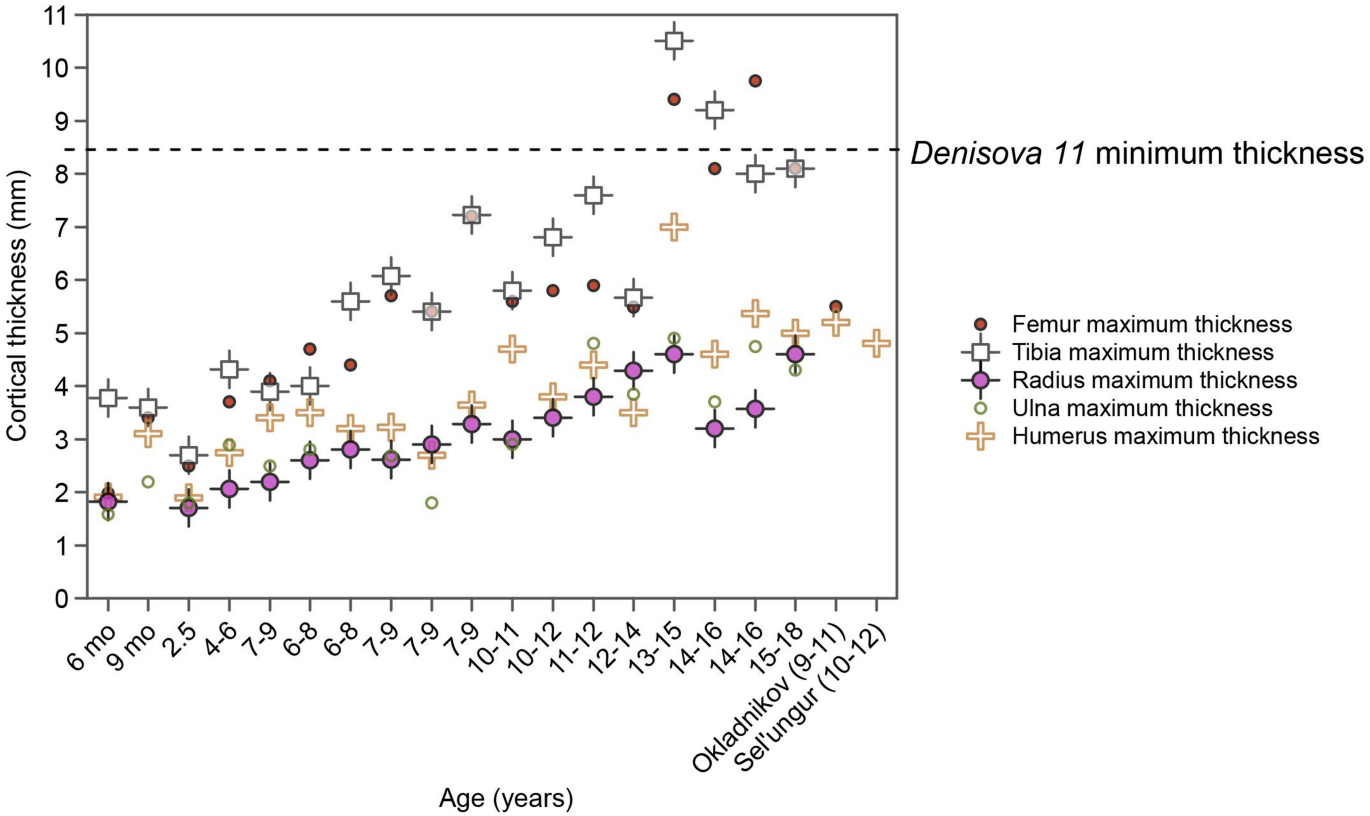
Sequencing and data processing. Sequencing was performed on Illumina platforms (MiSeq or HiSeq 2500) using 76-cycle paired-end runs adapted to double-indexed libraries[31]. Bases were called using Bustard (Illumina). Adaptor sequences were trimmed and overlapping paired-end reads were merged into single sequences using leeHom[36]. Demultiplexing was carried out using jivebunny[7]. Sequences generated from a given library were merged using SAMtools[37] and aligned to the human reference genome (hg19/GRCh37) with the decoy sequences as previously described[2] using BWA[38] with parameters adjusted to ancient DNA[6]. PCR duplicates were collapsed using bam-rmdup (https://bitbucket.org/ustenzel/biohazard) and DNA fragments of length ≥35 bases that mapped within regions of unique mappability (Map35_100% from a previous publication[8]) with a mapping quality of 25 or higher[7] were used for analyses. Further filtering criteria used for certain analyses are described in the Supplementary Information.

Reporting summary. Further information on research design is available in the Nature Research Reporting Summary linked to this paper.

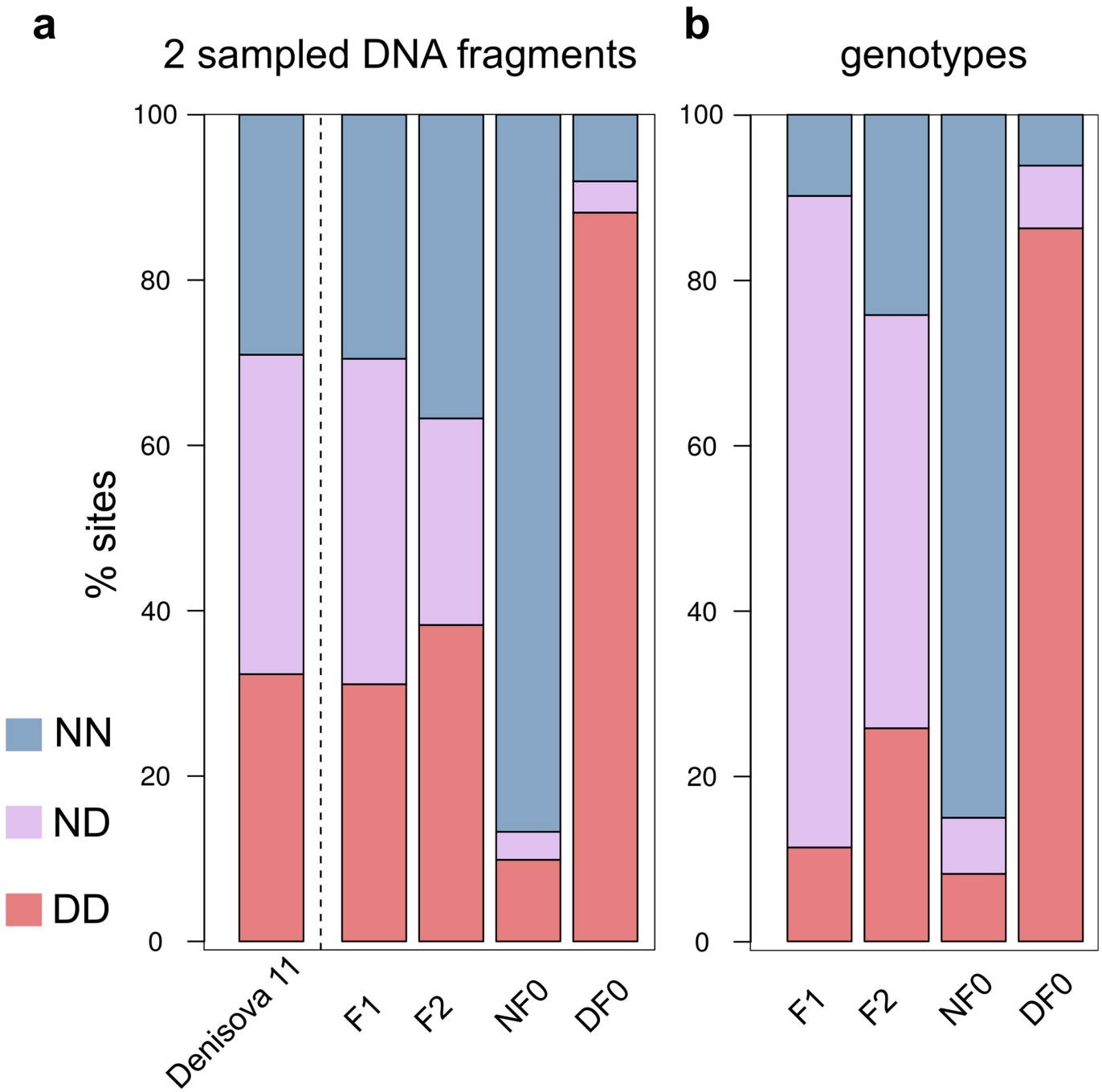
Code availability. The computer code used for simulations is included in Supplementary Information 6.

Data availability. Sequences generated from Denisova 11 have been deposited in the European Nucleotide Archive under study accession number PRJEB24663.

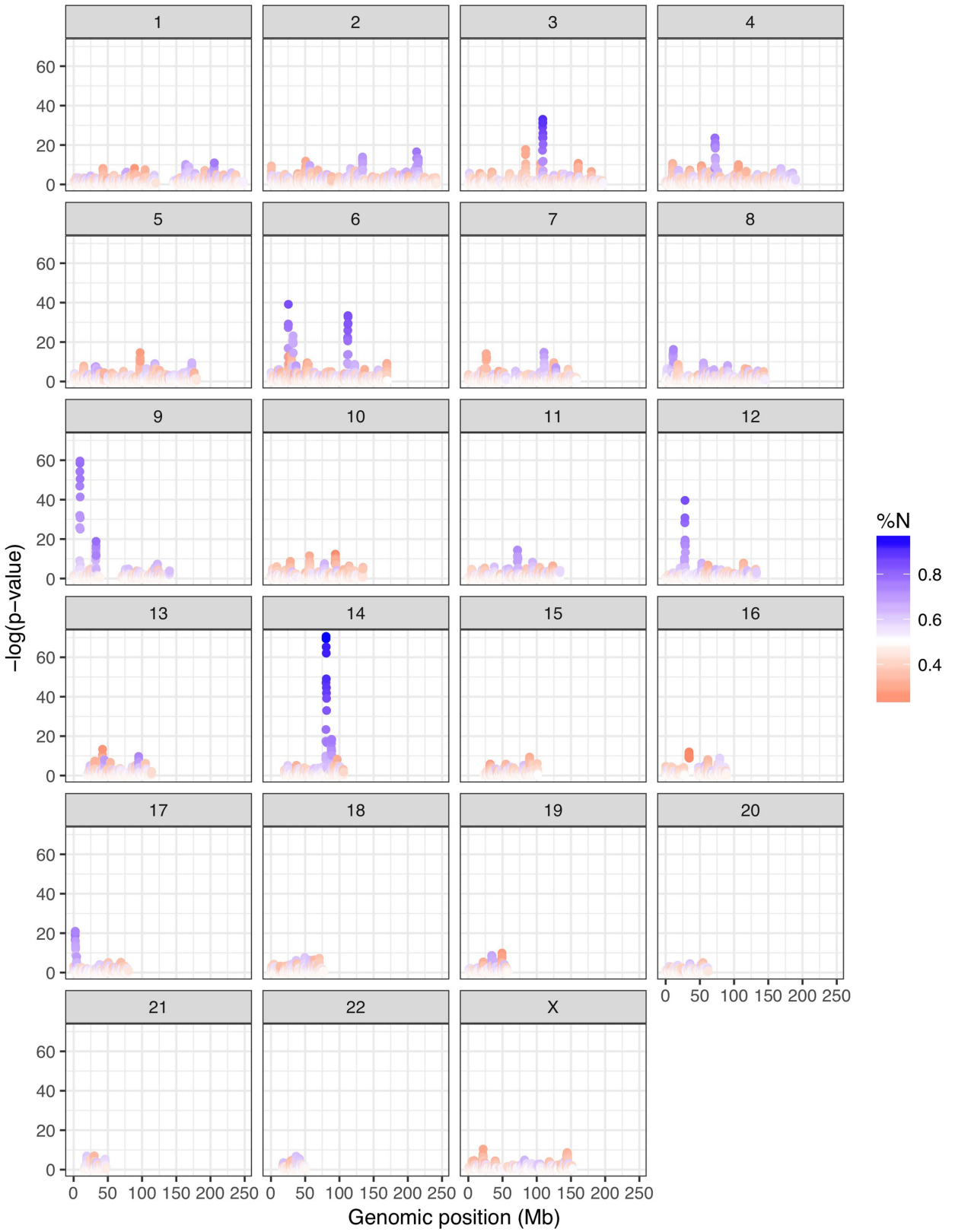
30. Slon, V. et al. Mammalian mitochondrial capture, a tool for rapid screening of DNA preservation in faunal and undiagnostic remains, and its application to Middle Pleistocene specimens from Qesem Cave (Israel). *Quat. Int.* **398**, 210–218 (2016).
31. Kircher, M., Sawyer, S. & Meyer, M. Double indexing overcomes inaccuracies in multiplex sequencing on the Illumina platform. *Nucleic Acids Res.* **40**, e3 (2012).
32. Dabney, J. & Meyer, M. Length and GC-biases during sequencing library amplification: a comparison of various polymerase-buffer systems with ancient and modern DNA sequencing libraries. *Biotechniques* **52**, 87–94 (2012).
33. DeAngelis, M. M., Wang, D. G. & Hawkins, T. L. Solid-phase reversible immobilization for the isolation of PCR products. *Nucleic Acids Res.* **23**, 4742–4743 (1995).
34. Slon, V. et al. Neandertal and Denisovan DNA from Pleistocene sediments. *Science* **356**, 605–608 (2017).
35. Meyer, M. & Kircher, M. Illumina sequencing library preparation for highly multiplexed target capture and sequencing. *Cold Spring Harb. Protoc.* https://doi.org/10.1101/pdb.prot5448 (2010).
36. Renaud, G., Stenzel, U. & Kelso, J. leeHom: adaptor trimming and merging for Illumina sequencing reads. *Nucleic Acids Res.* **42**, e141 (2014).
37. Li, H. et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
38. Li, H. & Durbin, R. Fast and accurate long-read alignment with Burrows–Wheeler transform. *Bioinformatics* **26**, 589–595 (2010).



Extended Data Fig. 1 | Comparison between cortical thickness of long bones from modern humans, Neanderthals and Denisova 11. Maximum cortical thickness of femora, tibiae, humeri, radii and ulnae from humans from the Bronze Age and two Neanderthals compared to the minimum thickness of Denisova 11 (dashed line).



Extended Data Fig. 2 | Comparison of the genome of Denisova 11 and simulated genomes. Percentage of sites at which Denisova 11 and genomes simulated under the demographic model described in Supplementary Information 6 carry two Neanderthal alleles (NN, blue), two Denisovan alleles (DD, red) or one allele of each type (ND, purple). **a**, Percentages calculated for two random DNA fragments from Denisova 11 and from simulated F_1, F_2, Neanderthal (NF_0) or Denisovan (DF_0) genomes. **b**, Proportions of sites for the simulated genotypes, before sampling two fragments.



Extended Data Fig. 3 | Neanderthal and Denisovan allele proportions from Denisova 11 in 1-Mb windows. The *y* axis shows −log(*P*) of the deviation of Neanderthal and Denisovan allele counts from the genome-wide average (χ^2 test of goodness-of-fit; see Supplementary Information 7). The colour shows the proportion of alleles matching the Neanderthal state (%N) within each 1-Mb window (100-kb steps, $n = 26{,}414$ windows).

Extended Data Table 1 | DNA extracts and DNA libraries prepared from the Denisova 11 specimen

	Extr. set	Bone powder [mg]	Pre-treatment [minutes]	Extract ID	Library prep. setup	Input in library [µl]	Molecules in library	Indexed library ID	DNA fragments sequenced	DNA fragments sequenced (L ≥35)	Mapped fragments (L ≥35, MQ ≥25)	Mapped fragments (%)	Unique fragments (L ≥35, MQ ≥25, Map35_100%)	Average fragment length [bp]	Fragments with C to T substitution
Denisova 11	1	27.4	15	E3259	A	10	2.27E+08	R5507	133,898,498	89,793,496	2,139,377	2.4	1,656,500	52.7	404,188
		27.8	15	E3261	A	10	2.00E+08	R5509	145,234,847	94,543,170	1,712,750	1.8	1,201,280	48.8	329,411
	2	29.0	15	E3652	B	10	4.33E+08 *	R5780	2,391,986	1,565,197	171,150	10.9	152,336	56.4	31,758
					C	5	4.63E+08	R9880	379,368,999	228,704,750	22,501,299	9.8	14,767,988	56.5	3,028,792
					C	5	4.03E+08	R9881	333,009,774	203,041,282	20,747,195	10.2	13,805,425	56.6	2,850,253
		29.7	15	E3654	B	10	4.16E+08 *	R5782	2,671,910	1,669,048	81,750	4.9	72,730	54.5	15,618
		33.5	15	E3655	B	10	3.49E+08 *	R5783	2,348,997	1,510,249	199,762	13.2	177,860	59.3	31,952
					C	5	4.19E+08	R9882	368,237,790	225,412,495	27,395,573	12.2	17,849,890	59.8	3,173,678
					C	5	3.69E+08	R9883	343,471,978	224,455,462	28,522,051	12.7	18,048,369	60.4	3,215,383
	3	27.1	30	E3922	C	10	7.43E+07	R9873	348,156,224	222,947,600	62,160,161	27.9	17,009,638	53.3	4,282,134
Controls	1	ENC	15	E3262	A	10	2.55E+07	R5510	12,220	4,123	38	0.9	35	55.2	2
		LNC	-	-	A	-	7.60E+06	R5521	12,444	2,170	11	0.5	10	47.2	1
	2	ENC	15	E3663	B	10	1.83E+06 *	R5791	32,008	4,183	473	11.3	412	51.2	8
		LNC	-	-	B	-	2.54E+06 *	R5792	31,455	2,908	70	2.4	58	49.0	3
	3	ENC	30	E3926	C	10	2.73E+07	R9877	61,825	13,861	2,472	17.8	2,145	57.0	9
		LNC	-	-	C	-	1.30E+07	R9888	68,130	5,275	100	1.9	67	46.2	6

Data are shown by DNA extraction set, and libraries prepared in the same setup are denoted with the same letter (A, B or C). Relevant negative controls are marked in grey. The number of molecules in each library was quantified by digital droplet PCR or quantitative PCR (denoted by asterisk). The numbers of DNA fragments sequenced per library are indicated for the combined data from all sequencing runs. Mapped fragments were counted if they were at least 35 bases long and mapped to the human reference genome with a mapping quality of 25 or higher; and their percentage was calculated out of sequenced fragments of length 35 bases or more. Following the removal of PCR duplicates, unique DNA fragments were retained if they mapped to the reference genome within the used mappability track. Such fragments were considered to contain a terminal cytosine (C) to thymine (T) substitution relative to the human reference genome if a putative cytosine deamination was within the first three or last three bases of the strand. bp, base pairs; ENC, extraction negative control; Extr., extraction; L, length; LNC, library preparation negative control; Map35_100%, previously published mappability track[8]; MQ, mapping quality; Prep., preparation.

LETTER

https://doi.org/10.1038/s41586-018-0452-0

Past experience shapes sexually dimorphic neuronal wiring through monoaminergic signalling

Emily A. Bayer[1] & Oliver Hobert[1]*

Differences between female and male brains exist across the animal kingdom and extend from molecular to anatomical features. Here we show that sexually dimorphic anatomy, gene expression and function in the nervous system can be modulated by past experiences. In the nematode *Caenorhabditis elegans*, sexual differentiation entails the sex-specific pruning of synaptic connections between neurons that are shared by both sexes, giving rise to sexually dimorphic circuits in adult animals[1]. We discovered that starvation during juvenile stages is memorized in males to suppress the emergence of sexually dimorphic synaptic connectivity. These circuit changes result in increased chemosensory responsiveness in adult males following juvenile starvation. We find that an octopamine-mediated starvation signal dampens the production of serotonin (5-HT) to convey the memory of starvation. Serotonin production is monitored by a 5-HT1A serotonin receptor homologue that acts cell-autonomously to promote the pruning of sexually dimorphic synaptic connectivity under well-fed conditions. Our studies demonstrate how life history shapes neurotransmitter production, synaptic connectivity and behavioural output in a sexually dimorphic circuit.

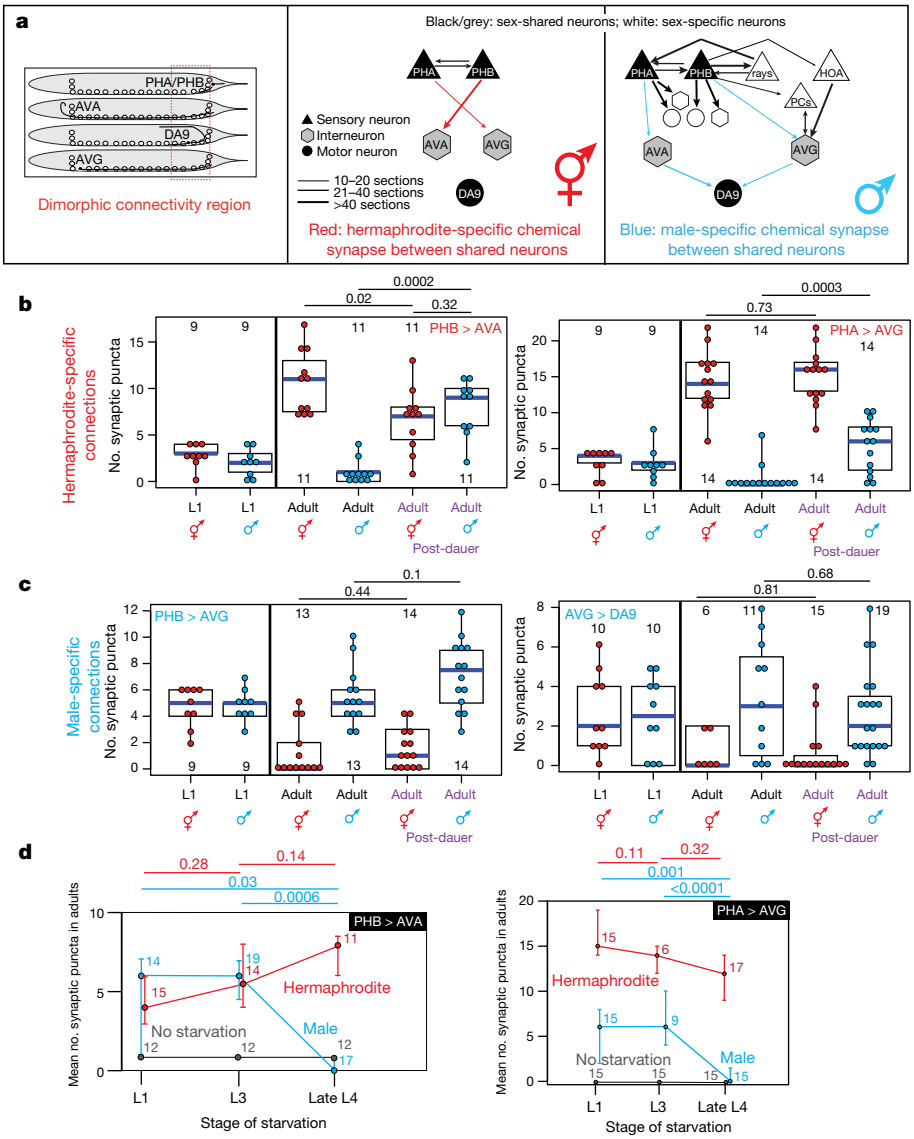


Fig. 1 | Starvation inhibits male-specific synaptic pruning. a, Schematic of adult chemical synaptic connectivity based on electron micrograph reconstruction[3,4]. Left, location of dimorphic synaptic connections within animal. mMN, male motor neuron; mIN, male interneuron. **b**, Hermaphrodite-specific PHB > AVA and PHA > AVG synaptic connections are maintained in post-dauer adult males as quantified by GFP reconstitution across synaptic partners (GRASP) and in vivo biotin labelling of intercellular contacts (iBLINC), respectively. Each dot represents one animal (red, hermaphrodite; cyan, male in all figures; n = number of animals, shown in each column), blue bars show median, black boxes represent quartiles and vertical black lines show range (**b**–**d**). Control animals are the progeny of the post-dauer animals. L1 animals (before synaptic pruning) shown to the left. P values shown above horizontal bars at top, calculated by two-sided Wilcoxon rank-sum test with Bonferroni corrections for multiple testing (see Methods) (**b**–**d**). Representative images shown (Extended Data Fig. 1a). While we observed a decrease in synaptic puncta in starved PHB > AVA hermaphrodites, this was not consistently reproducible across experimental replicates and thus is likely to represent an experimental artefact. **c**, The male-specific PHB > AVG and AVG > DA9 synaptic connections are pruned normally in post-dauer adult hermaphrodites. Representative images shown (Extended Data Fig. 1a). **d**, Starvation in the L1 or L3 stages, but not during L4, results in a failure of pruning of PHB > AVA and PHA > AVG in males. Animals were starved for 24 h (12 h of starvation was insufficient to affect male-specific pruning). Data shown as mean ± s.d. 'No starvation' shows progeny of starved animals as controls. n = number of animals, shown above each data point.

[1]Department of Biological Sciences, Howard Hughes Medical Institute, Columbia University, New York, NY, USA. *e-mail: or38@columbia.edu

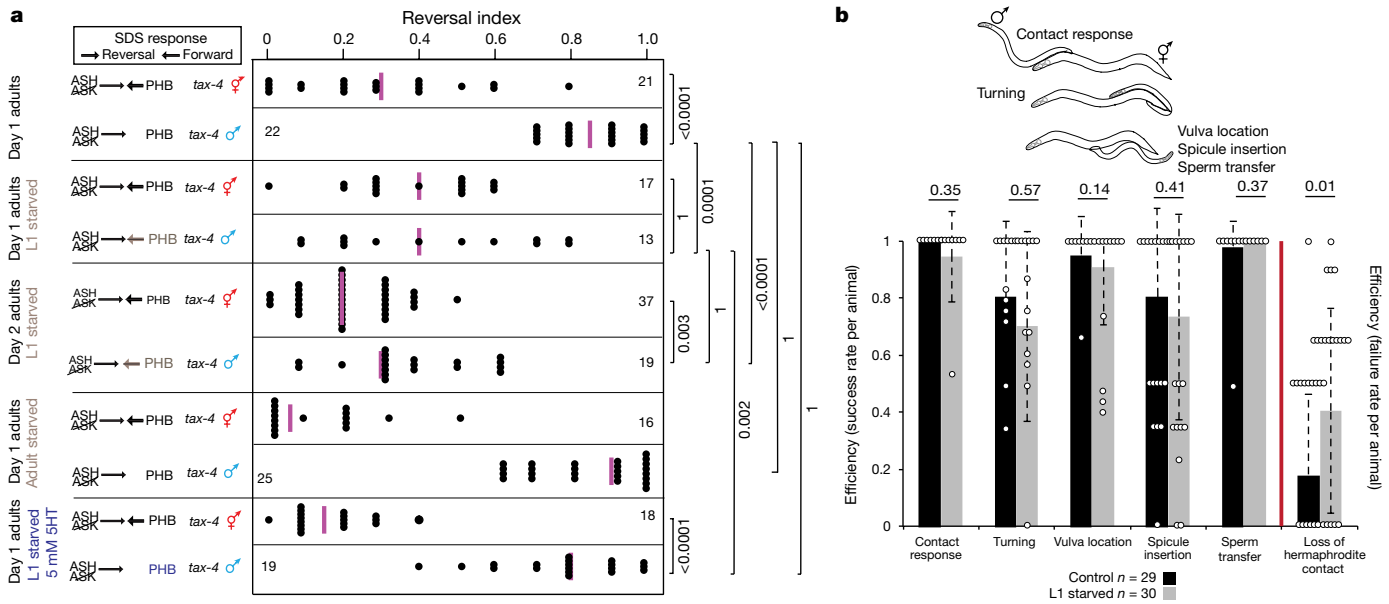


Fig. 2 | Juvenile starvation in males results in aberrant maintenance of juvenile behaviour. a, Males show enhanced chemosensory avoidance behaviour following L1 starvation. Left, predicted synaptic input into avoidance behaviour[1,5]. All experiments used *tax-4 (p678)* mutants to disable amphid input, as previously described[5]. Control animals are non-starved siblings of starved animals. Each dot represents the reversal index of one animal over ten experimental trials; vertical magenta bar is median. *n* = number of animals, shown in **a**, **b**. *P* values, two-sided Wilcoxon rank-sum test with Bonferroni corrections for multiple testing (see Methods) and 95% confidence intervals. Neither L1-starved nor continuously fed adult males change between adulthood days 1 and 2 (Extended Data Fig. 2a). **b**, Males show mating defects. Each step of mating (schematized on left) was scored in adult males following L1 starvation (grey) and compared to progeny of the starved generation (black). For behaviours to the left of the red bar, efficiency is shown as success rate per animal. For loss of hermaphrodite contact, efficiency is shown as failure rate per animal. Data shown as mean ± s.d. Each dot represents one animal (overlapping points graphically omitted). *P* values, two-sided Wilcoxon rank-sum test.

During sexual maturation, nervous systems develop a number of sexually dimorphic features. We investigated whether sexually dimorphic maturation of the nervous system is modulated by early-life experience in the two sexes of *C. elegans*, male and hermaphrodite[2]. Among the most notable sexual dimorphisms in *C. elegans* are sex-specific patterns of synaptic connectivity[3,4]. These sex-specific wiring differences arise during sexual maturation (the L4 stage) via sex-specific pruning of 'sex-hybrid' juvenile connectivity[1,4] (Fig. 1a). We found that early-life experience alters sex-specific synaptic pruning. Specifically, adult males that passed as sexually immature juveniles through a developmental arrest stage induced by unfavourable external conditions display no sex-specific pruning of the normally hermaphrodite-specific PHB > AVA and PHA > AVG synaptic connections (Fig. 1b, Extended Data Fig. 1a, b). By contrast, normally male-specific connections are pruned in hermaphrodites that passed as juveniles through this dauer stage (Fig. 1c). These observations suggest a sexually dimorphic sensitivity to the memory of environmental stress. While a number of stressors had no effect on male-specific synapse pruning (Extended Data Fig. 1c), starvation of males (but not hermaphrodites) at early juvenile stages fully recapitulated the effect of dauer passage on male-specific synapse pruning (Fig. 1d, Extended Data Fig. 1d). By contrast, starvation during the L4 stage did not result in defects in male-specific synaptic pruning (Fig. 1d). Synaptic pruning is first observable in the L4 stage[1], demonstrating that starvation prevents pruning if the stress occurs before its onset, but does not halt or reverse synaptic pruning.

The PHB phasmid sensory neurons modulate the avoidance response to noxious chemicals (for example, sodium dodecyl sulfate (SDS)) via synapses onto the AVA command interneurons in juvenile animals of both sexes and in adult hermaphrodites; in adult males, pruning of the PHB > AVA synaptic contacts eliminates this behavioural output[1,5]. Juvenile-starved adult males retain the ability to respond to SDS, as predicted by the failure to prune PHB > AVA (Fig. 2a). By contrast, males starved during adulthood (after pruning has already occurred) do not show any change in SDS avoidance behaviour (Fig. 2a), consistent with normal PHB > AVA pruning. Thus, a starvation experience allows males to retain juvenile sensory acuity and results in the loss of an adult behavioural sexual dimorphism. However, maintenance of juvenile sensory acuity affects other behaviours in the adult male. Adult males differ from adult hermaphrodites in that they search for, and mate with, hermaphrodites[6,7]. Following juvenile starvation, males display a defect in maintaining contact with hermaphrodites during mating, a phenotype previously associated with ablation of AVA (the command interneuron postsynaptic to both PHA and PHB)[8] (Fig. 2b, Extended Data Fig. 2b). Overall, adverse juvenile experience prevents male-specific synaptic pruning and increases noxious sensory responsiveness, but also decreases male mating efficiency.

In *C. elegans*, several feeding-dependent behaviours, such as locomotion and aversive memory, are known to be modulated by the monoamines serotonin and octopamine: serotonin signals well-fed conditions (similar to vertebrates) and octopamine (similar to norepinephrine in vertebrates[9]) signals starvation[10,11]. To investigate whether monoamine signalling regulates male-specific synaptic pruning, we supplemented well-fed animals with exogenous octopamine during the L1 stage and found that this mimicked starvation, suppressing male-specific synaptic pruning of the PHB > AVA and PHA > AVG connections (Fig. 3a, Extended Data Fig. 3a). Conversely, starving L1 animals in the presence of exogenous serotonin showed normal male pruning of the PHB > AVA and PHA > AVG connections, as well as retaining the SDS avoidance behavioural dimorphism (Figs. 2a, 3a). Consistent with this result, pruning was also rescued in males that were starved in the presence of the selective serotonin reuptake inhibitor fluoxetine or in *mod-5*/SERT mutants, which show increased extracellular serotonin[12] (Fig. 3b, Extended Data Fig. 3b, c). No pruning defects were observed in males defective for dopamine production (Extended Data Fig. 3d).

The effects of exogenous octopamine and serotonin on pruning suggest that monoamine signalling is involved in regulating male-specific synaptic pruning, but do not address the requirement for endogenous octopamine and serotonin. Therefore, we measured the expression of the rate-limiting enzymes of octopamine and serotonin synthesis. In larval *C. elegans*, octopamine is produced exclusively in the sex-shared RIC interneuron class by tyramine-β-hydroxylase (encoded by *tbh-1*)[13]. Consistent with our results with exogenous octopamine treatment,

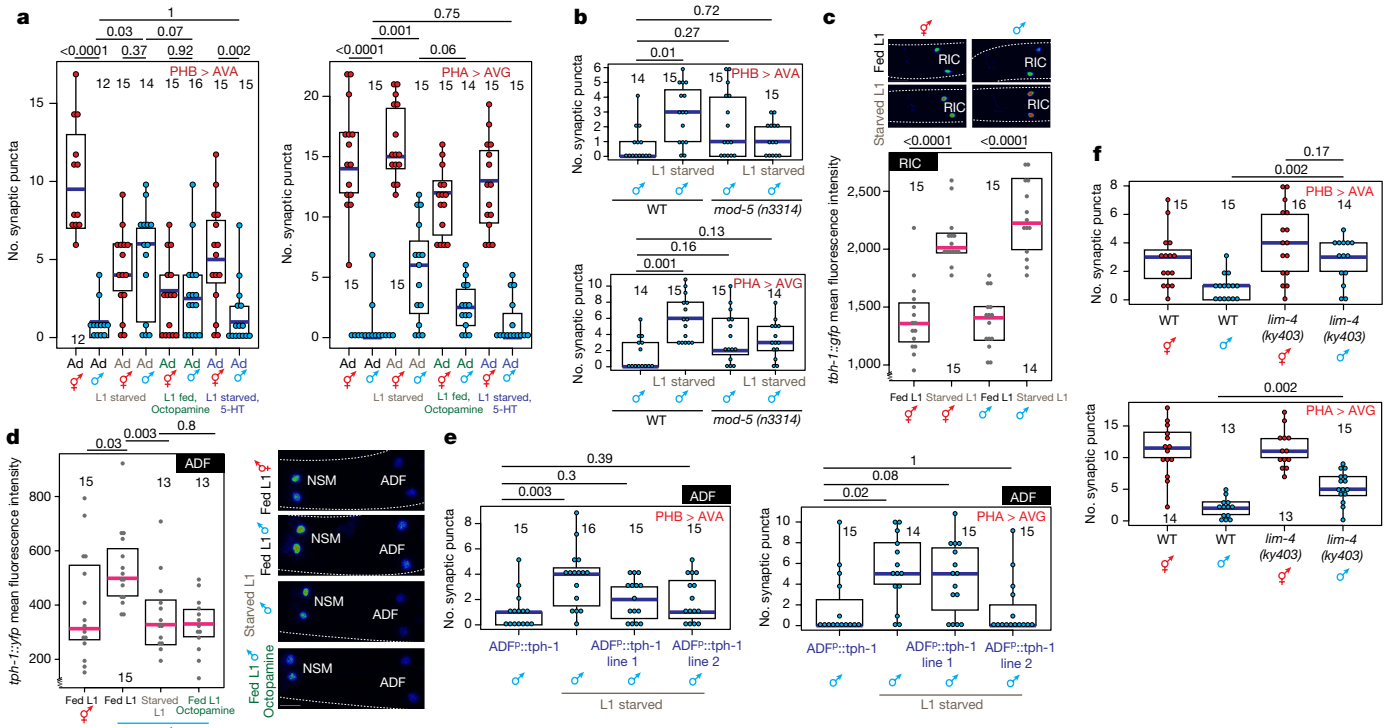


Fig. 3 | Serotonin and octopamine convey feeding and starvation signals via the ADF neurons. a, Octopamine and 5-HT mimic starvation and feeding, respectively. Quantification of PHB > AVA and PHA > AVG synaptic connectivity in adults of both sexes using GRASP or iBLINC (**a**, **b**, **e**, **f**). Each dot represents the number of synaptic puncta in one animal; blue bars, median, black boxes, quartiles; vertical black lines, range (**a**, **b**, **e**, **f**). *n* = number of animals, shown in each column (**a**–**f**). *P* values, two-sided Wilcoxon rank-sum test with Bonferroni corrections (see Methods) for multiple testing (**a**–**f**). Representative images shown (Extended Data Fig. 3a). **b**, *mod-5* mutations rescue pruning defects in PHB > AVA and PHA > AVG connections in L1-starved adult males. Representative images shown (Extended Data Fig. 3c). **c**, Starvation induces octopamine production. Expression of a *tbh-1* reporter in RIC in fed or starved L1 animals. Heat-map rendered fluorescence intensity images above, quantification below. Scale bars, 10 μm. Magenta bar, median; black boxes, quartiles (**c**, **d**). Anterior left, dorsal up in all figures. **d**, *tph-1* transcription in ADF neurons is increased in males and decreased by starvation or exogenous octopamine. Expression of a *tph-1* transcriptional fosmid in ADF in fed L1 animals, starved L1 animals or L1 animals fed in the presence of exogenous octopamine. **e**, Overexpression of *tph-1* in ADF mimics the rescuing effect of exogenous 5-HT. Quantification of PHB > AVA and PHA > AVG synaptic connectivity using GRASP and iBLINC. Two independent transgenic lines were tested for each experiment. L1-starved animals without transgenic lines are siblings of transgenic animals and controls are non-starved adult males with transgenic arrays. Representative images shown (Extended Data Fig. 6c). **f**, Genetic ablation of ADF results in failure to prune. Quantification of PHB > AVA and PHA > AVG synaptic connectivity in both sexes in *lim-4* mutants[24].

starvation at either the adult[14] or L1 stage (Fig. 3c) increases transcription of *tbh-1* (Fig. 3c, Extended Data Fig. 4a). The expression of the serotonin-synthesizing enzyme tryptophan hydroxylase (encoded by *tph-1*) is also dependent on starvation, at least in adult animals[15], with an octopamine receptor conveying the starvation signal[16]. At the first larval stage, *tph-1* is expressed exclusively in two sex-shared classes of serotonin-producing neurons, the bilateral NSM and ADF neuron pairs[15]. Under well-fed conditions, a *tph-1* fosmid reporter was dimorphically expressed in both NSM and ADF, with L1 males showing higher expression than hermaphrodites (Fig. 3d, Extended Data Fig. 4b). Upon starvation during L1, *tph-1* expression decreases acutely in male ADF neurons, but not NSM neurons (Fig. 3d, Extended Data Fig. 4b). Addition of exogenous octopamine to otherwise well-fed animals mimicked the effect of starvation on *tph-1* transcription in male ADF neurons (Fig. 3d), whereas the addition of tyramine did not (Extended Data Fig. 4c). Confirming the effects observed with a *tph-1* reporter transgene, single-molecule fluorescent in situ hybridization (smFISH) of endogenous *tph-1* mRNA also revealed sexually dimorphic expression in ADF neurons, as well as a transcriptional decrease (in ADF, but not NSM) following starvation in L1 (Extended Data Fig. 4d).

Unlike the upregulation of the octopamine-producing *tbh-1* gene, downregulation of *tph-1* reporter expression persisted even after animals were returned to food (Extended Data Fig. 4a, e). Furthermore, in continuously fed animals, *tph-1* expression peaks during the L3 stage in both sexes (the onset of sexual maturation), whereas *tbh-1* expression is unchanged over the course of development (Extended Data Fig. 4a, e). Notably, this upregulation never occurs in L1-starved animals, and this starvation memory is dependent on *tbh-1* and, thus, on octopamine production (Extended Data Fig. 5a). We also observed decreased *tph-1* transcription in ADF but not NSM during the L3 stage in SER-6 octopamine receptor mutants, corroborating the relevance of octopamine signalling for regulating serotonin levels during sexual maturation (Extended Data Fig. 5b). Conversely, exogenous serotonin did not suppress the upregulation of *tbh-1* (Extended Data Fig. 5c). Demonstrating the functional relevance of persistent serotonin downregulation, male-specific pruning defects in L1-starved animals were rescued by supplementation of exogenous serotonin during the L3 stage (after the animals had been returned to food; Extended Data Fig. 6a). We conclude that a temporary increase in starvation-induced octopamine production triggers a long-term alteration in serotonin production.

To further investigate whether levels of serotonin production in the sex-shared ADF neuron regulate sex-specific synaptic pruning during male sexual maturation, we also used neuron-specific promoters to overexpress *tph-1* in NSM or ADF during L1 starvation. Overexpression of *tph-1* in ADF, but not in NSM, rescued sex-specific pruning of PHB > AVA and PHA > AVG synapses in males that had been starved (Fig. 3e, Extended Data Fig. 6b, c). Through selective removal of serotonin production from either NSM or ADF neurons (using cell-specific *tph-1* mutant rescue experiments and genetic elimination of either NSM or ADF; Fig. 3f, Extended Data Fig. 6d, e), we found that serotonin production specifically in ADF is not only sufficient but also required to modulate neuronal male-specific synaptic maturation. This analysis also revealed that NSM-secreted serotonin has an earlier role in phasmid neuron migration and morphology (Extended Data Fig. 6d).

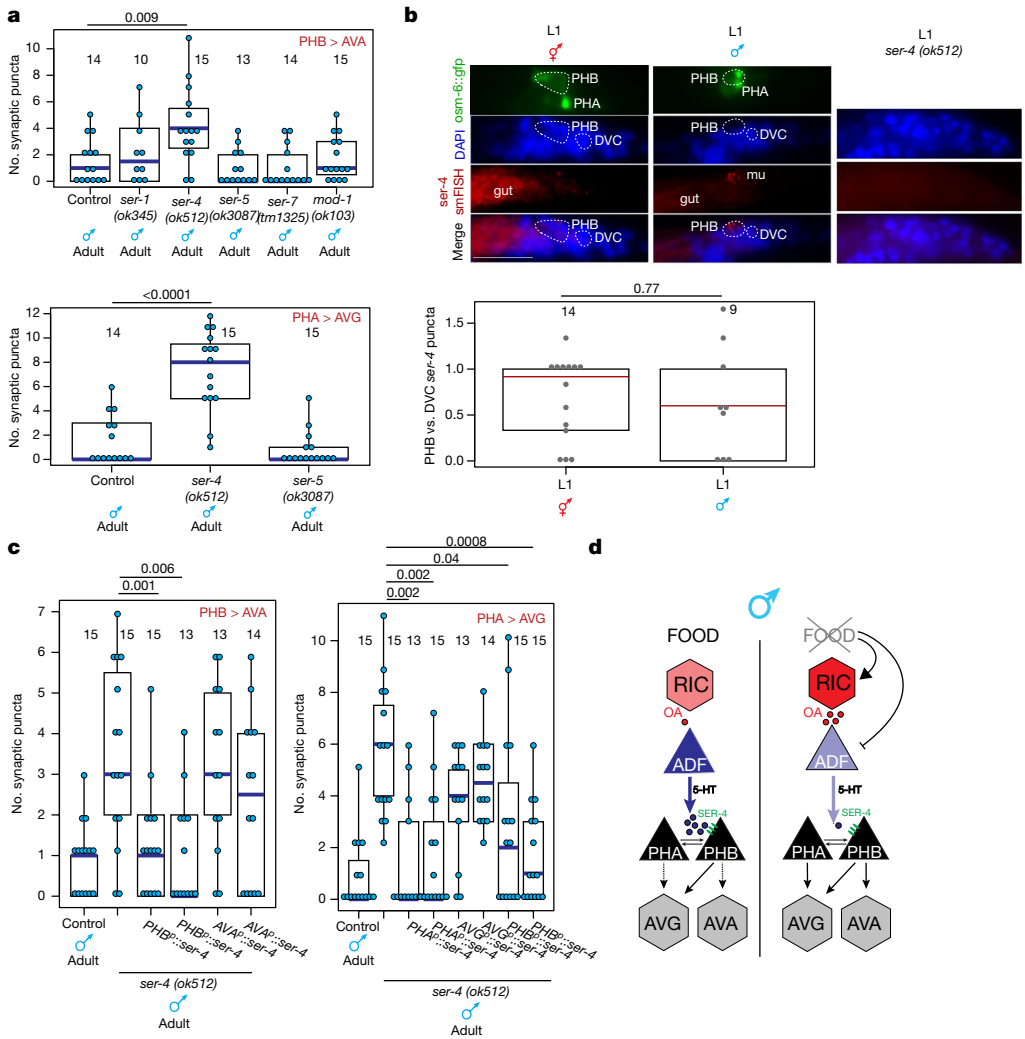


Fig. 4 | The *ser-4*/5-HT1A serotonin receptor acts downstream of the feeding signal. a, Quantification of PHB > AVA and PHA > AVG synaptic connectivity in adult males mutant for serotonin receptors. Each dot represents one animal, blue bar represents median, black box represents quartiles and vertical black bars represent range (**a**, **c**). *n* = number of animals, shown in each column (**a**–**c**). *P* values, two-sided Wilcoxon rank-sum test with Bonferroni corrections for multiple testing (see Methods) (**a**–**c**). **b**, *ser-4* smFISH puncta are present in PHB in both sexes at L1. Maximum intensity projections of one half of each animal are shown for *ser-4* smFISH and DAPI panels; the centre slice from the stacks is shown in the GFP panel. Merge shows overlay of *ser-4* smFISH onto DAPI. Each grey dot represents the normalized number of *ser-4* smFISH puncta in one PHB neuron, red bar indicates median and box indicates quartiles. Higher-magnification individual *z*-slices are shown in Extended Data Fig. 7b. **c**, Expression of *ser-4* cDNA in PHB or PHA rescues the PHB > AVA and PHA > AVG pruning defects, respectively. Two independent transgenic lines were evaluated for each promoter. **d**, Summary of effects of feeding (left) and starvation (right).

To investigate how serotonin signals from the ADF neurons in the head to the phasmid–interneuron connections in the tail, we analysed PHA > AVG and/or PHB > AVA connectivity in males mutant for each of the four metabotropic serotonin receptors in *C. elegans*[17,18]. In well-fed males lacking the *ser-4*/5-HT1A receptor, juvenile synaptic connectivity was maintained, phenocopying juvenile starvation (Fig. 4a, Extended Data Fig. 7a). This phenotype was neither enhanced by juvenile starvation nor rescued by exogenous serotonin in *ser-4* mutants, suggesting that the *ser-4* serotonin receptor acts directly and non-redundantly downstream of the feeding cue (Extended Data Fig. 8a). Furthermore, *tph-1* transcription levels were unaffected in the *ser-4* mutant, confirming that *ser-4* does not feed back onto serotonin production itself (Extended Data Fig. 8b).

In both *C. elegans* and vertebrates, *ser-4*/5-HT1A responds to extrasynaptic serotonergic signalling[19,20], raising the possibility that serotonin from ADF neurons might directly modulate phasmid connectivity. Using smFISH probes against *ser-4* mRNA, we identified *ser-4* transcripts in several neurons, including PHB but not PHA neurons, and detected no signals in *ser-4* mutants (confirming probe specificity) (Fig. 4b, Extended Data Fig. 7b). To determine whether *ser-4* exerts its role neuron-autonomously within the phasmid sensory circuit, we performed cell-specific rescue experiments. We found that AVA-driven *ser-4* was unable to rescue PHB > AVA pruning in *ser-4* mutants, but that expression of *ser-4* in PHB rescued the pruning defect, consistent with a cell-autonomous function (Fig. 4c, Extended Data Fig. 7a).

Serotonin signalling can also act non-cell-autonomously in the phasmids to modulate synaptic pruning. Specifically, we found that PHB-expressed *ser-4* rescued pruning of the PHA > AVG connection (Fig. 4c, Extended Data Fig. 7a). Moreover, although we found that *ser-4* was expressed in the PHB but not PHA neurons, expression of *ser-4* in PHA rescued the PHA > AVG pruning defect (Fig. 4c, Extended Data Fig. 7a). The extensive, male-specific gap junctions that connect the PHA and PHB neurons[3,4] may allow PHA and PHB to readily cross-communicate a signal from a serotonin receptor.

Early-life stress is known to have long-lasting effects in vertebrates[21], and this also involves serotonin signalling[22]. For example, prenatal stress in mouse models of human serotonin transporter variants affects adult memory, anxiety and depressive-like behaviour[23].

Yet although both behavioural effects and some molecular effects of early-life stress in vertebrates have been identified, it has been difficult to link specific molecular changes to corresponding behavioural outcomes, a notion further confounded by complex contributions of genetic background[22]. Here we show that a juvenile starvation stress results in lasting circuit and behavioural effects in adult male, but not hermaphrodite, *C. elegans* by affecting serotonin levels during sexual maturation (Fig. 4d). We find that serotonin normally signals extrasynaptically to act as a cue for male-specific synaptic pruning. These results provide insight into both how temporary early-life stress can result in specific lasting changes to the nervous system and how stress can intersect with sexual maturation and result in differential effects between the two sexes, a feature that has also been observed in vertebrate models[21,22]. We anticipate that still-unknown genetic components underlying the many differences in sexual maturation between the two sexes of both *C. elegans* and vertebrates will shed light on the sexually differential sensitivity to juvenile starvation and other early-life stress.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, statements of data availability and associated accession codes are available at https://doi.org/10.1038/s41586-018-0452-0.

Received: 18 January; Accepted: 10 July 2018;
Published online 27 August 2018.

1. Oren-Suissa, M., Bayer, E. A. & Hobert, O. Sex-specific pruning of neuronal synapses in *Caenorhabditis elegans*. *Nature* **533**, 206–211 (2016).
2. Barr, M. M., García, L. R. & Portman, D. S. Sexual dimorphism and sex differences in *Caenorhabditis elegans* neuronal development and behavior. *Genetics* **208**, 909–935 (2018).
3. White, J. G., Southgate, E., Thomson, J. N. & Brenner, S. The structure of the nervous system of the nematode *Caenorhabditis elegans*. *Phil. Trans. R. Soc. Lond. B* **314**, 1–340 (1986).
4. Jarrell, T. A. et al. The connectome of a decision-making neural network. *Science* **337**, 437–444 (2012).
5. Hilliard, M. A., Bargmann, C. I. & Bazzicalupo, P. C. *elegans* responds to chemical repellents by integrating sensory inputs from the head and the tail. *Curr. Biol.* **12**, 730–734 (2002).
6. Lipton, J., Kleemann, G., Ghosh, R., Lints, R. & Emmons, S. W. Mate searching in *Caenorhabditis elegans*: a genetic model for sex drive in a simple invertebrate. *J. Neurosci.* **24**, 7427–7434 (2004).
7. Liu, K. S. & Sternberg, P. W. Sensory regulation of male mating behavior in *Caenorhabditis elegans*. *Neuron* **14**, 79–89 (1995).
8. Sherlekar, A. L. et al. The *C. elegans* male exercises directional control during mating through cholinergic regulation of sex-shared command interneurons. *PLoS ONE* **8**, e60597 (2013).
9. Goldstein, J. L. et al. Surviving starvation: essential role of the ghrelin-growth hormone axis. *Cold Spring Harb. Symp. Quant. Biol.* **76**, 121–127 (2011).
10. Churgin, M. A., McCloskey, R. J., Peters, E. & Fang-Yen, C. Antagonistic serotonergic and octopaminergic neural circuits mediate food-dependent locomotory behavior in *Caenorhabditis elegans*. *J. Neurosci.* **37**, 7811–7823 (2017).
11. Harris, G. et al. The monoaminergic modulation of sensory-mediated aversive responses in *Caenorhabditis elegans* requires glutamatergic/peptidergic cotransmission. *J. Neurosci.* **30**, 7889–7899 (2010).
12. Jafari, G., Xie, Y., Kullyev, A., Liang, B. & Sze, J. Y. Regulation of extrasynaptic 5-HT by serotonin reuptake transporter function in 5-HT-absorbing neurons underscores adaptation behavior in *Caenorhabditis elegans*. *J. Neurosci.* **31**, 8948–8957 (2011).
13. Alkema, M. J., Hunter-Ensor, M., Ringstad, N. & Horvitz, H. R. Tyramine functions independently of octopamine in the *Caenorhabditis elegans* nervous system. *Neuron* **46**, 247–260 (2005).
14. Tao, J., Ma, Y. C., Yang, Z. S., Zou, C. G. & Zhang, K. Q. Octopamine connects nutrient cues to lipid metabolism upon nutrient deprivation. *Sci. Adv.* **2**, e1501372 (2016).
15. Liang, B., Moussaif, M., Kuan, C. J., Gargus, J. J. & Sze, J. Y. Serotonin targets the DAF-16/FOXO signaling pathway to modulate stress responses. *Cell Metab.* **4**, 429–440 (2006).
16. Noble, T., Stieglitz, J. & Srinivasan, S. An integrated serotonin and octopamine neuronal circuit directs the release of an endocrine signal to control *C. elegans* body fat. *Cell Metab.* **18**, 672–684 (2013).
17. Carre-Pierrat, M. et al. Characterization of the *Caenorhabditis elegans* G protein-coupled serotonin receptors. *Invert. Neurosci.* **6**, 189–205 (2006).
18. Harris, G. P. et al. Three distinct amine receptors operating at different levels within the locomotory circuit are each essential for the serotonergic modulation of chemosensation in *Caenorhabditis elegans*. *J. Neurosci.* **29**, 1446–1456 (2009).
19. Gürel, G., Gustafson, M. A., Pepper, J. S., Horvitz, H. R. & Koelle, M. R. Receptors and other signaling proteins required for serotonin control of locomotion in *Caenorhabditis elegans*. *Genetics* **192**, 1359–1371 (2012).
20. Riad, M. et al. Somatodendritic localization of 5-HT1A and preterminal axonal localization of 5-HT1B serotonin receptors in adult rat brain. *J. Comp. Neurol.* **417**, 181–194 (2000).
21. Lajud, N. & Torner, L. Early life stress and hippocampal neurogenesis in the neonate: sexual dimorphism, long term consequences and possible mediators. *Front. Mol. Neurosci.* **8**, 3 (2015).
22. Houwing, D. J., Buwalda, B., van der Zee, E. A., de Boer, S. F. & Olivier, J. D. A. The serotonin transporter and early life stress: translational perspectives. *Front. Cell. Neurosci.* **11**, 117 (2017).
23. van den Hove, D. L. et al. Differential effects of prenatal stress in 5-Htt deficient mice: towards molecular mechanisms of gene × environment interactions. *PLoS ONE* **6**, e22715 (2011).
24. Zheng, X., Chung, S., Tanabe, T. & Sze, J. Y. Cell-type specific regulation of serotonergic identity by the *C. elegans* LIM-homeodomain factor LIM-4. *Dev. Biol.* **286**, 618–628 (2005).

Acknowledgements We thank Q. Chen for generating transgenic strains, K. Ashrafi for providing DNA constructs, the CGC (supported by the NIH P40 OD010440) for strains and members of the Hobert laboratory, I. Greenwald, and M. Oren-Suissa for comments on the manuscript. This work was supported by the HHMI and NIH (R37NS039996, O.H., F31NS096863, E.A.B.).

Author contributions E.A.B and O.H. designed the experiments and wrote the manuscript. E.A.B. performed the experiments.

Competing interests The authors declare no competing interests.

Additional information
Extended data is available for this paper at https://doi.org/10.1038/s41586-018-0452-0.
Supplementary information is available for this paper at https://doi.org/10.1038/s41586-018-0452-0.
Reprints and permissions information is available at http://www.nature.com/reprints.
Correspondence and requests for materials should be addressed to O.H.
Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

METHODS

Strains. Wild-type strains were *C. elegans* variety Bristol, strain N2. Worms were maintained by standard methods[25]. Worms were grown at 20 °C on nematode growth media (NGM) plates seeded with bacteria (*Escherichia coli* OP50) as a food source. GRASP and iBLINC reagents have been previously described[1,26,27]. A detailed list of all mutant and transgenic strains used is available in Supplementary Table 1.

Cloning and constructs. To generate *pEAB42* (*srg-13p::ser-4::SL2::tagRFP*), *ser-4* cDNA was amplified off *pEAB69* (*ser-4* cDNA in pUC57) and inserted into *pEAB6* (*srg-13p::fem-3::SL2::tagRFP*) using restriction-free cloning to replace the *fem-3* cDNA. To generate *pEAB43* (*gpa-6p::ser-4::SL2::tagRFP*), *pEAB59* (*flp-18p::ser-4::SL2::tagRFP*), and *pEAB60* (*inx-18p::ser-4::SL2::tagRFP*), the *srg-13* promoter was digested out of the SphI–XmaI sites of *pEAB42* and each promoter was ligated in (from *pEAB1* (2.2 kb upstream of *srg-13* fused to *wCherry*), *pEAB3* (2.6 kb of the *gpa-6* promoter fused to *GFP*), *pEAB10* (3.1 kb of the *flp-18* promoter fused to *GFP*), and *pMO10* (intron 2 of *inx-18* with the AIY enhancer site deleted[1] fused to *wCherry*).

pKA805 (*psrh-142::tph-1::GFP*) and *pKA807* (*pceh-2::tph-1::GFP*) were gifts from K. Ashrafi.

Microscopy. Worms were anaesthetized using 100 mM sodium azide (NaN_3) and mounted on 5% agar on glass slides. Worms were analysed with Nomarski optics and fluorescence microscopy, using a Zeiss 880 confocal laser-scanning microscope. Multidimensional data were reconstructed as maximum intensity projections using Zeiss Zen software. For GRASP experiments, animals were imaged using a 63× objective and puncta were quantified by scanning the original full *Z*-stack for distinct dots in the area where the processes of the two neurons overlap. GRASP experiments were scored blinded to genotype for mutant analysis and rescue array analysis, and experimental condition for starvation, heat stress, hyperosmotic stress and serotonin rescue experiments. For fluorescence intensity experiments, animals were imaged using a 40× objective with fixed imaging settings and quantification was performed using the Zeiss Zen software by measuring the mean fluorescent intensity of the neuronal cell body in the centre *Z*-slice of each neuron, and then averaging the intensity of the left and right neurons of each pair to control for differences based on *Z*-position. The sex of L1 animals was determined using rectal epithelial cell morphology and/or coelomocyte position.

Figure preparation. Plots for GRASP and expression data were generated in R using the beeswarm package. Figures were prepared using Adobe Photoshop CS6 and Adobe Illustrator CS6.

Statistics and reproducibility. Two-tailed Wilcoxon rank-sum tests were performed in R in addition to post hoc Bonferroni corrections to adjust *P* values for number of pairwise tests in all cases where more than two pairwise statistical tests were performed. Freeman–Halton extension of one-sided Fisher exact tests were performed for the categorical data in Extended Data Fig. 6d. All experiments were repeated independently (technical replicates) a minimum of twice with similar results; any case for which replication failed is indicated in the corresponding figure legend. Within each figure, all datasets contain only representative biological replicates. No statistical method was used to determine sample size and animals were not randomized into trial groups.

Starvation and neurotransmitter assays. L1 starvation assays were performed by plating a synchronized population of embryos (following hypochlorite treatment of gravid adults) onto unseeded NGM and allowing 12 h at 20 °C for embryo hatching, followed by 24 h at 20 °C for starvation (unless otherwise noted) before transfer onto seeded NGM by washing L1 animals using M9 buffer. Later starvation assays (L3, L4, adult) were performed by washing synchronized populations grown at 20 °C off seeded NGM plates using M9 buffer, performing at least 3 washes in M9 at 510rcf to remove OP50, and then plating animals on NGM plates without bacto-peptone to prevent the growth of any residual OP50 for 24 h before transfer back to seeded NGM.

For neurotransmitter assays, all drugs were mixed into NGM medium (for serotonin starvation assays, without bacto-peptone) before pouring to a final concentration of 20 mg/ml (octopamine), 5 mM (serotonin), 5 mM (tyramine) or 0.1 mg/ml (fluoxetine). Drugs used were 5-hydroxytryptamine hydrochloride (Sigma-Aldrich catalogue #H9523), (1)-octopamine hydrochloride (Sigma-Aldrich catalogue #O0250), tyramine hydrochloride (Sigma-Aldrich catalogue #T2879), and fluoxetine hydrochloride, USP (Spectrum Chemical catalogue # F1200). L1 animals were synchronized onto drug-containing plates identically to starvation assays (above) and then transferred onto plain seeded NGM plates 24 h post-hatching.

L1 heat stress assay. Heat stress assays were performed by transferring synchronized, well-fed L1 animals (6 h post-hatching) to a 35 °C incubator for 30 min, and then recovering plates at 20 °C until adulthood.

L1 osmotic stress assay. Hyperosmotic stress assays were performed by growing L1 animals on plates containing 200 mM NaCl for 24 h after hatching (in the continuous presence of food) and then transferring animals to standard NGM plates until adulthood.

SDS-avoidance behaviour. The SDS avoidance assay was based on procedures as described[5]. A small drop of solution containing either the repellent (0.1% SDS in M13 buffer) or buffer (M13 buffer: 30 mM Tris-HCl pH 7.0, 100 mM NaCl, 10 mM KCl) is delivered near the tail of an animal while it moves forward. Once in contact with the tail, the drop surrounds the entire animal by capillary action and reaches the anterior amphid sensory organs. The drop was delivered using 10-μl glass calibrated pipets (VWR international) pulled by hand on a flame to reduce the diameter of the tip. The capillary pipette was mounted in a holder with rubber tubing and operated by mouth. Assayed worms were transferred individually to fresh non-wet unseeded NGM plates. Each assay started by testing the animals with drops of M13 buffer alone. The response to each drop was scored as reversing or not reversing. The avoidance index is the number of reversal responses divided by the total number of trials. An interstimulus interval of at least two minutes was used between successive drops to the same animal. Each animal was tested ten times, and all animals that survived all ten trials were included in the datasets. Two biological and technical replicates were performed for each experiment. Each replicate began with $n = 40$ animals for each condition.

Mate-searching behaviour. The male-leaving assay was based on procedures as described[6]. Males were separated from hermaphrodites at the L4 stage and transferred to assay plates 24 h later. Assays were performed on NGM plates (5 cm diameter) and seeded with 20 μl OP50 to create a small lawn that was allowed to grow overnight. Distance from the lawn was recorded every 3 h over a total assay period of 12 h, and then mean distance from the lawn was calculated for each assayed male. Two biological and technical replicates were performed for each experiment.

Mating behaviour assays. Mating assays were based on procedures as described[7]. Males were picked at the L4 stage and kept apart from hermaphrodites for 24 h, either following 24 h of starvation during L1 or not. One male was transferred to a plate covered with a thin fresh OP50 lawn containing 10–15 adult *unc-31 (e928)* hermaphrodites. These hermaphrodites move very little, allowing easy recording of male behaviour. Hermaphrodites were also isolated from the opposite sex at the L4 stage and used 24 h later, and were always well-fed. Animals were monitored and the sequence of events was recorded within a 10 min window or until the male ejaculated, whichever occurred first. Males were digitally recorded using the Exo Labs model 1 camera mounted on Nikon Eclipse E400 compound microscope with long-distance ×20 lenses. Per cent efficiency (per mating behaviour step) = 100 × (number of successful performances of mating behaviour step/number of times the male attempted the mating behaviour step). Per cent loss of hermaphrodite contact = 100 × (number of times male lost contact with hermaphrodite and eventually re-initiated the mating sequence/ number of hermaphrodites the male contacted during assay). Two biological and technical replicates were performed for each experiment.

Single molecule fluorescent in situ hybridization (smFISH). smFISH was performed as described[28]. In brief, L1 larvae (6 h post-hatching on food for control or 24 h starved for L1 starved) were washed off NGM and fixed (4% PFA) for 45 min at room temperature. Fixed worms were washed twice with PBS, resuspended in 70% ethanol, and incubated over two nights at 4 °C. The fixed sample was centrifuged and incubated with wash buffer for 5 min. After wash buffer was removed, hybridization buffer containing the *ser-4* or *tph-1* and *ric-4* probes (designed using the Stellaris RNA FISH probe, *ser-4* and *tph-1* conjugated to CalFluor 590, *ric-4* conjugated to Quasar 670, from Biosearch Technologies) was added to the sample. The sample was incubated overnight at 37 °C. The sample was subsequently incubated in wash buffer with DAPI for 30 min at 37 °C protected from light. The sample was suspended in 2× SSC and then resuspended in GLOX buffer for 2 min. The sample was then resuspended in GLOX buffer to which glucose oxidase was added. The sample was then mounted and imaged immediately. Images were acquired using an automated fluorescence microscope (Zeiss, AXIO Imager Z.2) with a 63× objective. Acquisition of *Z*-stack images (each slice 0.3 μm thick) was performed with ZEN 2 pro software. Representative images are shown following max-projection of *Z*-stacks in Zeiss Zen software. Puncta were quantified by scanning through the *Z*-slices sequentially in Zeiss Zen software. Each staining experiment was performed with two biological and technical replicates.

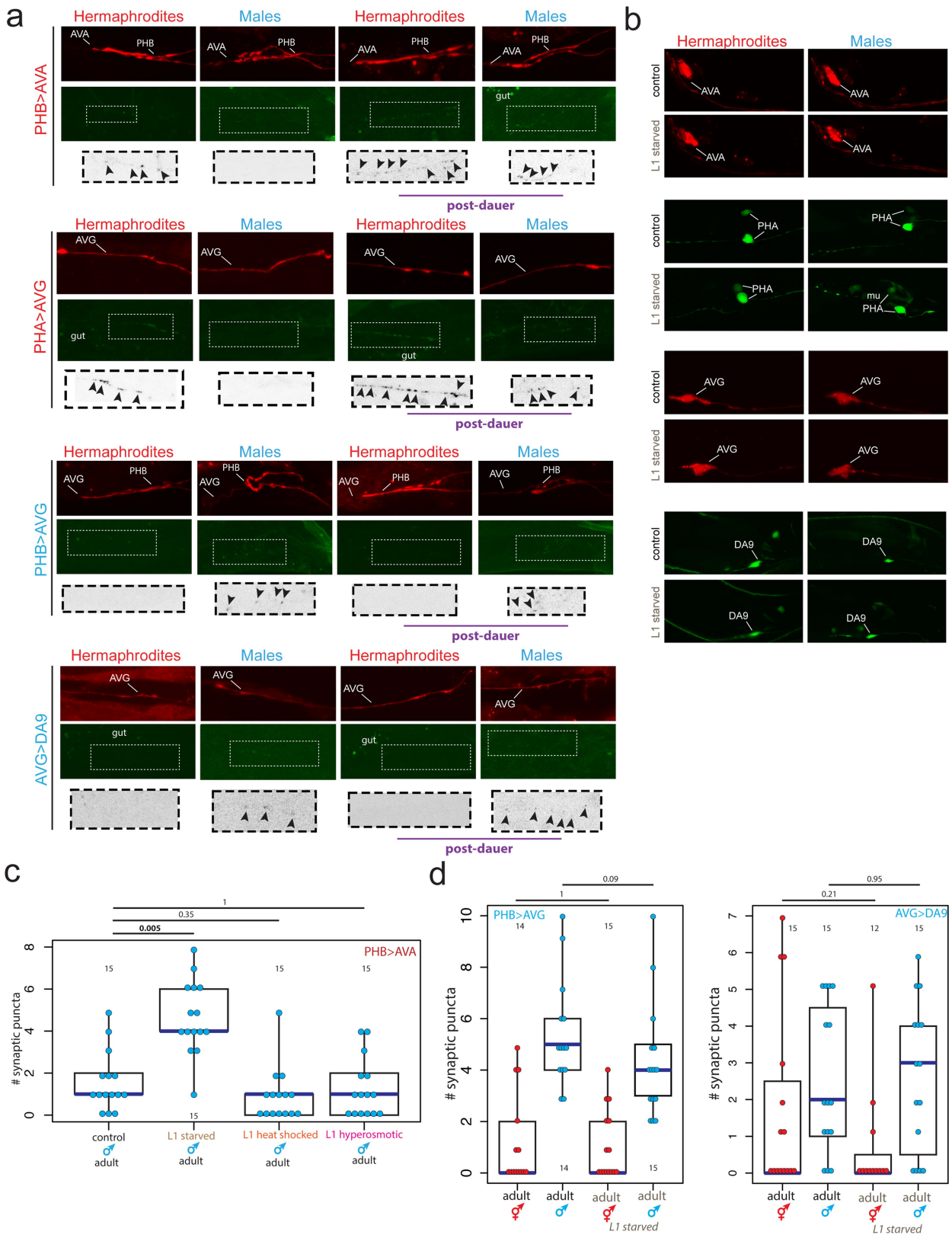
For *ser-4* smFISH, *osm-6::gfp* (*oyIs59*) labels all ciliated sensory neurons and was used to identify the phasmid neurons. Identification of DVC and anal depressor muscle (mu) was based on a previously published *ser-4* transcriptional reporter[19]. The number of smFISH puncta in one PHB neuron was normalized to the number of puncta in DVC (in the same animal) to control for staining fluctuations.

For *tph-1* smFISH, NSM and ADF neurons were identified by their stereotyped position, and smFISH puncta colocalizing with the DAPI-stained nuclei were quantified.

Reporting summary. Further information on experimental design is available in the Nature Research Reporting Summary linked to this paper.

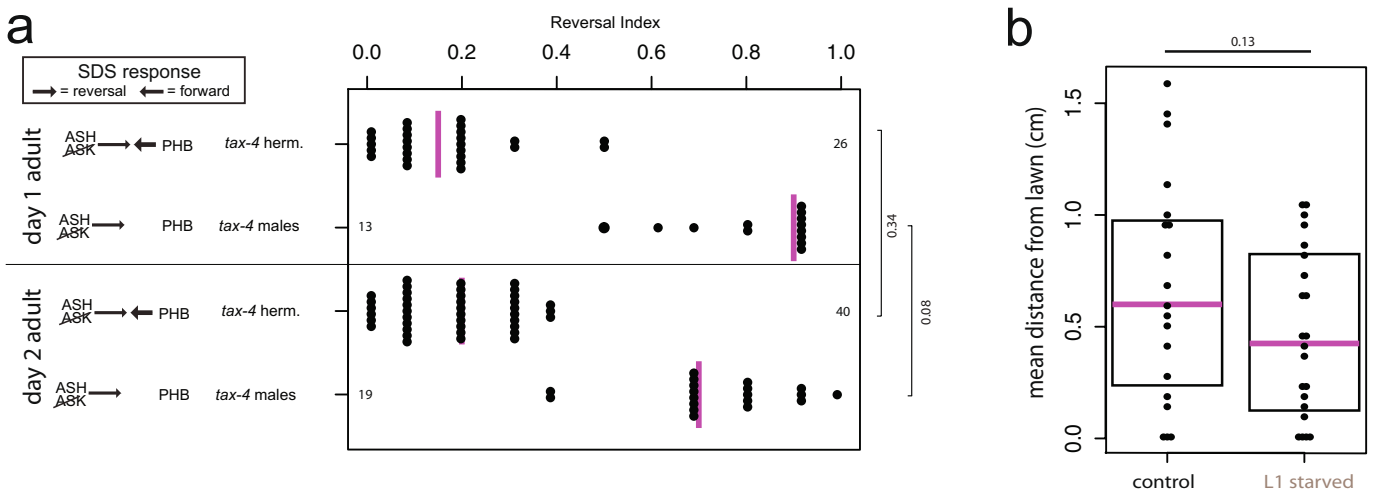
Data availability. The data that support the findings of this study are available from the corresponding author upon reasonable request.

25. Brenner, S. The genetics of *Caenorhabditis elegans*. *Genetics* **77**, 71–94 (1974).
26. Feinberg, E. H. et al. GFP reconstitution across synaptic partners (GRASP) defines cell contacts and synapses in living nervous systems. *Neuron* **57**, 353–363 (2008).
27. Desbois, M., Cook, S. J., Emmons, S. W. & Bülow, H. E. Directional *trans*-synaptic labeling of specific neuronal connections in live animals. *Genetics* **200**, 697–705 (2015).
28. Ji, N. & van Oudenaarden, A. Single molecule fluorescent in situ hybridization (smFISH) of *C. elegans* worms and embryos. *WormBook* https://doi.org/10.1895/wormbook.1.153.1 (2012).
29. Zhang, F. et al. The LIM and POU homeobox genes ttx-3 and unc-86 act as terminal selectors in distinct cholinergic and serotonergic neuron types. *Development* **141**, 422–435 (2014).

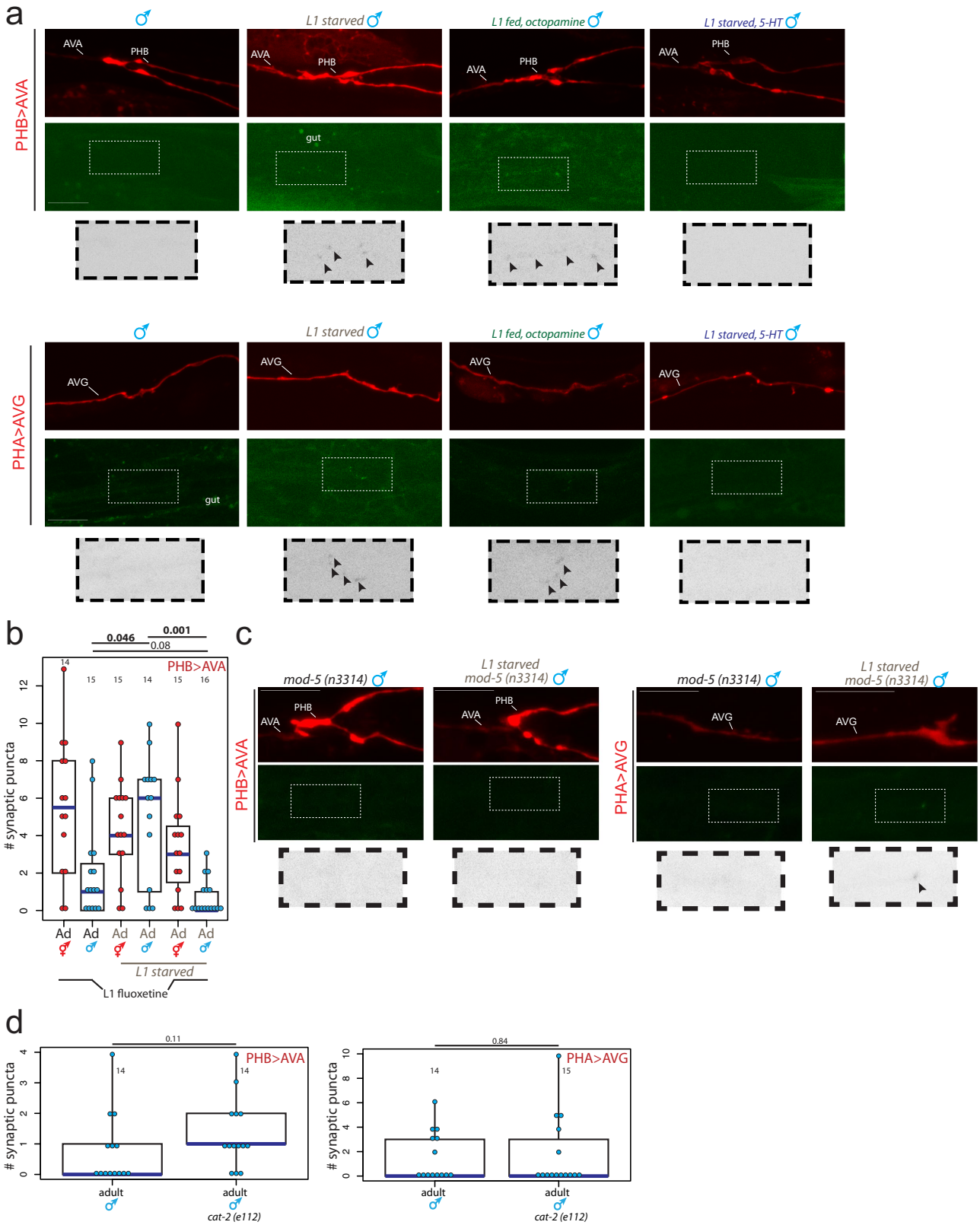


Extended Data Fig. 1 | See next page for caption.

Extended Data Fig. 1 | Trans-synaptic labelling by GRASP. a, The normally hermaphrodite-specific PHB > AVA and PHA > AVG synaptic connections fail to prune in post-dauer adult males, and the normally male-specific PHB > AVG and AVG > DA9 connections are pruned in post-dauer adult hermaphrodites. Top, red cytoplasmic axon label; middle, GRASP (PHB > AVA, PHB > AVG, AVG > DA9) or iBLINC signal (PHA > AVG); bottom, magnified inset of colour-inverted synaptic puncta with arrowheads to indicate puncta. Intestinal auto-fluorescence is labelled 'gut'. Representative images shown; for quantification and replication, see Fig. 1b, c and Methods. **b**, Starvation does not affect expression of the cell-specific promoters used for GRASP, and thus the effects on synaptic pruning are not an artefact of changes in promoter expression. Representative maximum intensity projection images of control animals and animals recovered from 24 h of L1 starvation are shown. mu, muscle (*srg-13p* is variably expressed in some muscle cells in addition to PHA). **c**, Neither L1 heat shock (30 min at 35 °C) nor L1 osmotic stress (24 h on plates with 200mM NaCl) affects male-specific pruning of the PHB > AVA connection. Each dot represents one animal (n = number of animals, shown in each column), blue bars show median, black boxes represent quartiles and vertical black lines show range. **c**, **d**, P values calculated by two-sided Wilcoxon rank-sum test with Bonferroni corrections for multiple testing (where applicable; see Methods). **d**, L1 starvation does not affect the normally male-specific PHB > AVG and AVG > DA9 synapses. Control animals are the progeny of starved animals.

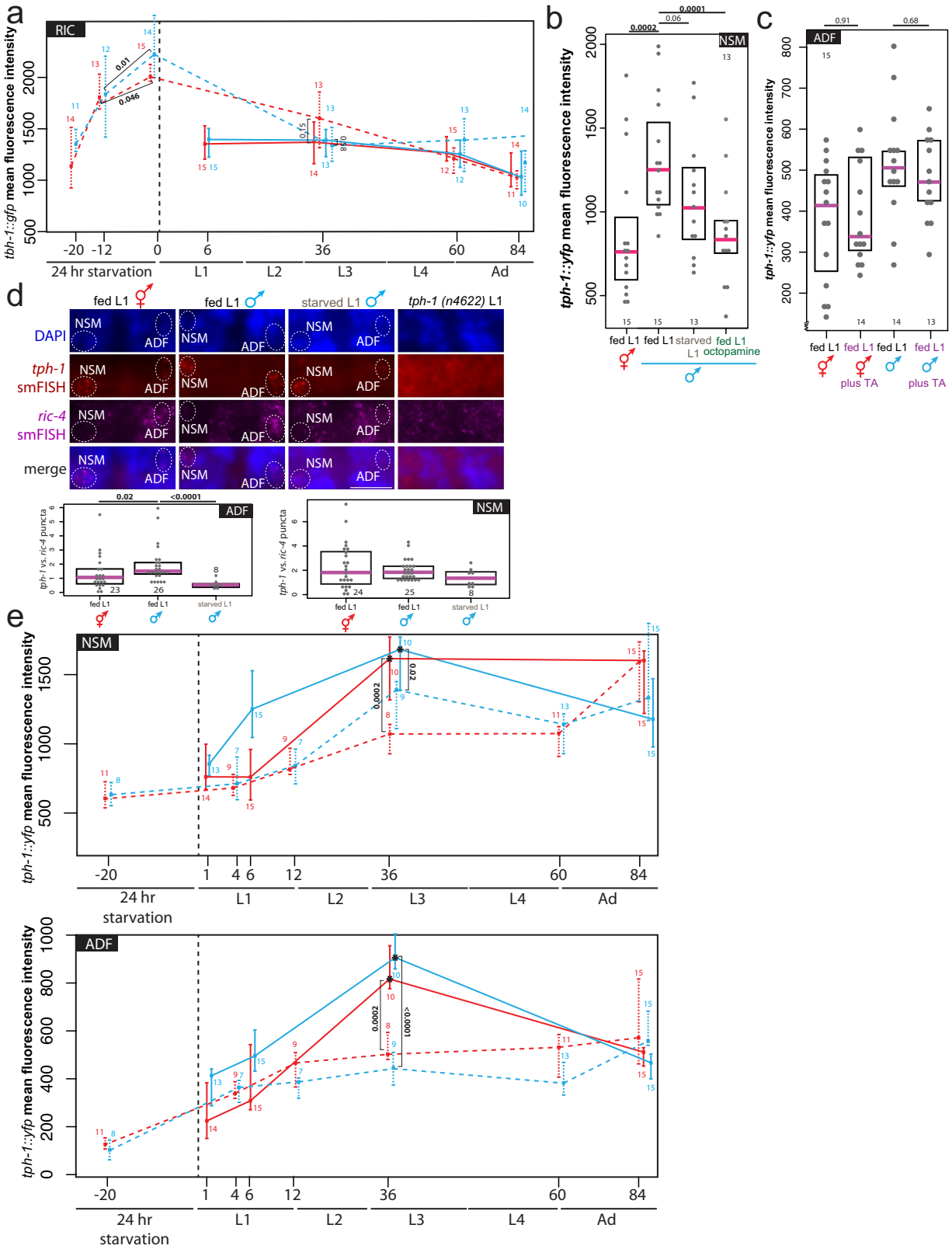


Extended Data Fig. 2 | Behaviours unaffected by starvation or additional adult development. a, SDS avoidance response is unchanged in day 2 adults. Left, predicted synaptic input into avoidance behaviour by relevant amphid and phasmid neurons[1,5]. Each dot represents the average reversal index of one animal over ten experimental trials, median shown with vertical magenta bar. *P* values calculated by two-sided Wilcoxon rank-sum test. **b**, Male mate-searching behaviour is unaffected by L1 starvation. Each dot represents the average distance one male travelled away from a bacterial lawn at four time points over 12 h ($n = 48$ animals control, 20 animals L1 starved), in the absence of hermaphrodites. Magenta bars indicate median, black boxes indicate quartiles. *P* values calculated by two-tailed *t*-test. By contrast, we did find mate-searching defects in adult males following recovery from dauer, suggesting that these males may have additional changes to the nervous system (data not shown).



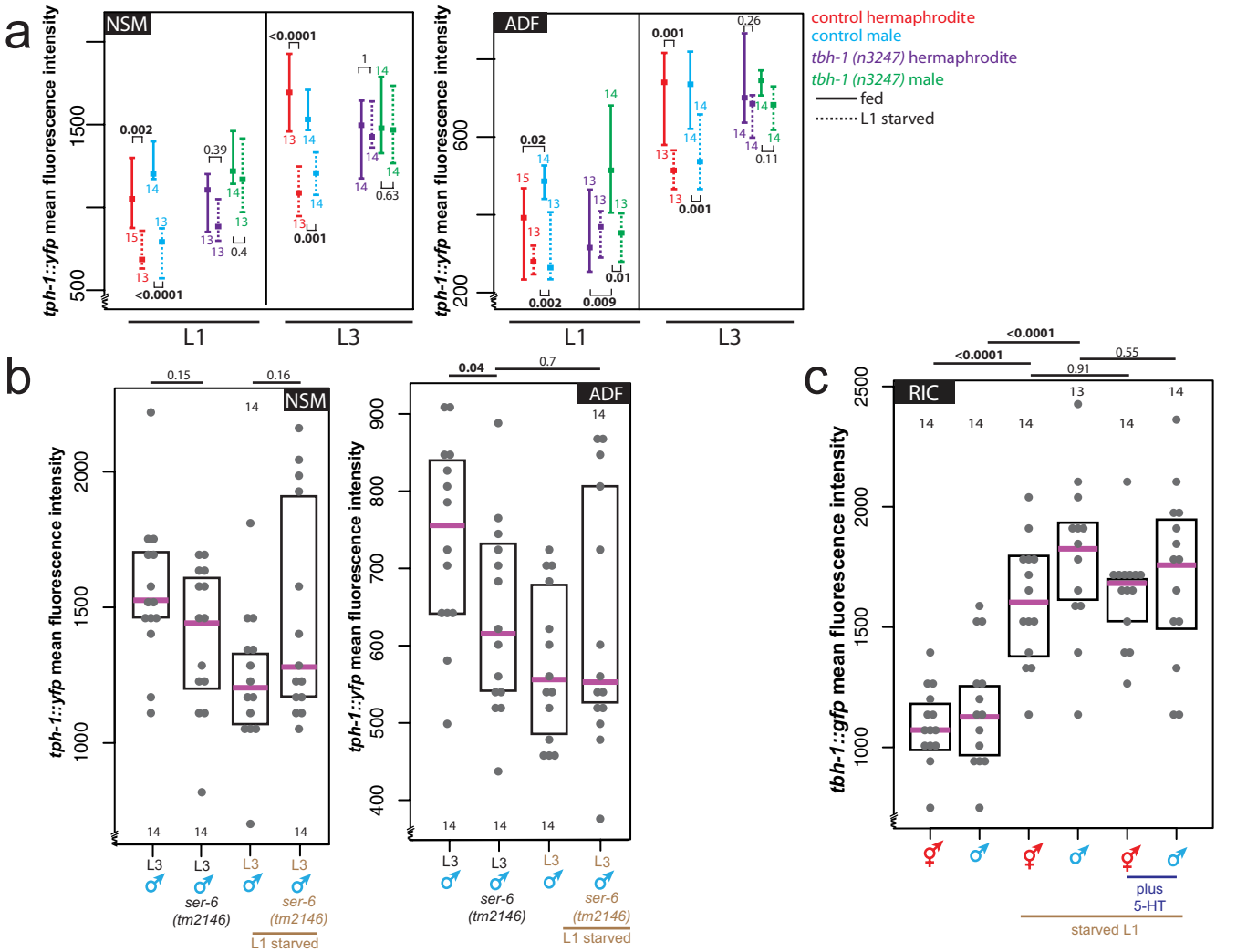
Extended Data Fig. 3 | See next page for caption.

Extended Data Fig. 3 | Effects of starvation and exogenous or endogenous monoamine signalling on synaptic connectivity. **a**, The normally hermaphrodite-specific PHB > AVA and PHA > AVG synaptic connections fail to prune in adult males following L1 starvation or treatment with exogenous octopamine during L1, but can be rescued by exogenous serotonin during L1 starvation. Top, red cytoplasmic axon label; middle, GRASP (PHB > AVA) or iBLINC signal (PHA > AVG); bottom, magnified inset of colour-inverted synaptic puncta with arrowheads to indicate puncta. Scale bars, 10 μm, all panels. Representative images shown; for quantification and replication, see Fig. 3a and Methods. **b**, Increases in extrasynaptic 5-HT through fluoxetine suppresses the failure to prune synapses. Quantification of PHB > AVA synaptic connectivity in adults following exposure to fluoxetine during L1 (on food), L1 starvation without fluoxetine, or L1 starvation in the presence of exogenous fluoxetine (0.1 mg ml^{-1}). Each dot represents one animal (n = number of animals, shown in each column), blue bars show median, black boxes represent quartiles, vertical black lines show range (**b**, **d**). P values calculated by two-sided Wilcoxon rank-sum test with Bonferroni corrections for multiple testing (where applicable; see Methods). **c**, The effect of L1 starvation on male-specific synaptic pruning is rescued in the *mod-5* mutant background. Representative images shown; for quantification and replication, see Fig. 3b and Methods. **d**, Loss of dopamine production (in a mutant for the *cat-2* tyrosine hydroxylase) has no effect on the pruning of the PHB > AVA and PHA > AVG connections in males.

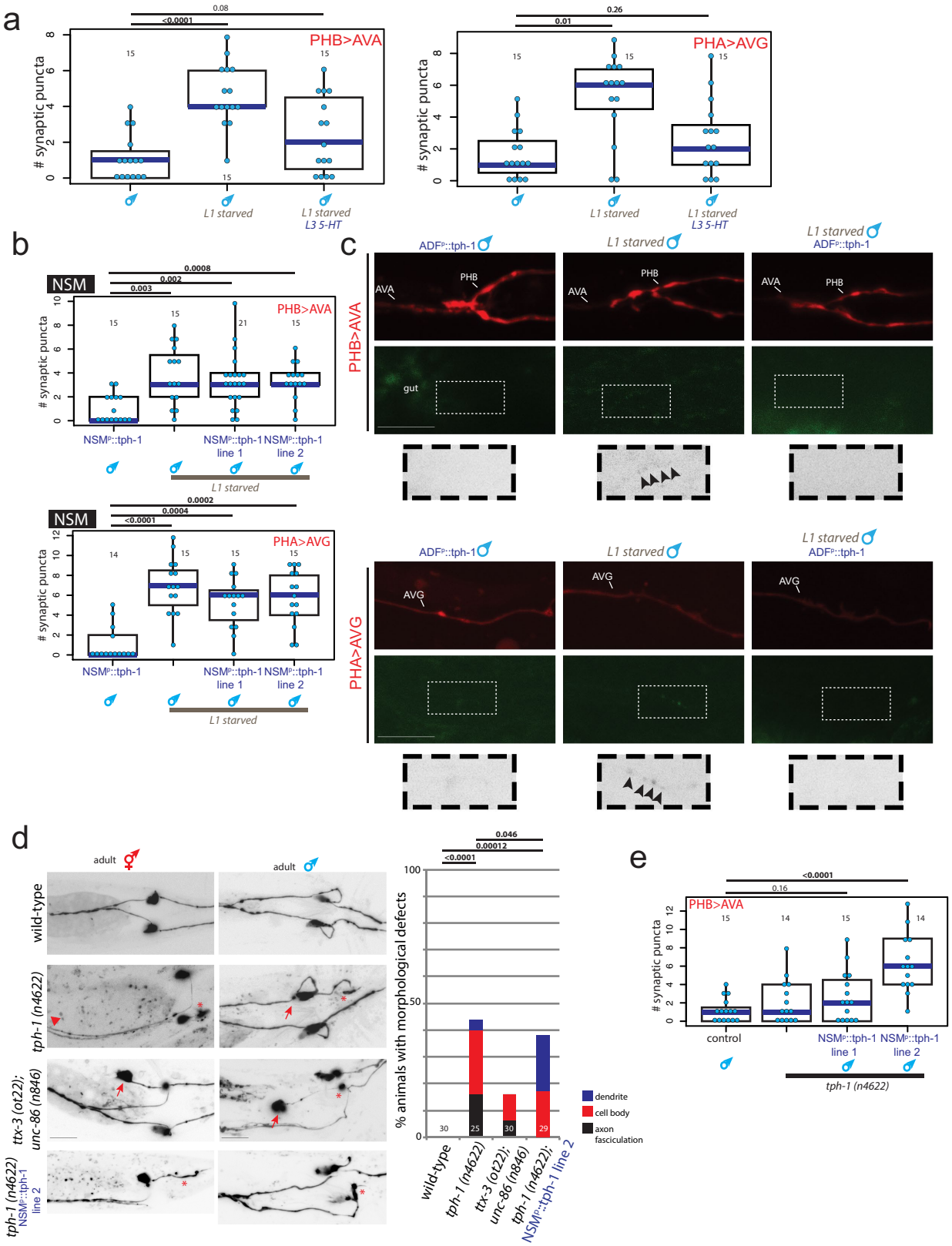


Extended Data Fig. 4 | See next page for caption.

Extended Data Fig. 4 | Effects of starvation on transcription of *tbh-1* and *tph-1*. **a**, Time-course of *tbh-1* transcriptional levels in fed (solid lines) and L1-starved (dashed lines) animals. *tbh-1* levels are even higher after 24 h of starvation than after 12 h of starvation, providing a molecular correlate for our observation that 12 h of starvation is insufficient to affect male-specific synaptic pruning (Fig. 1d). Larval stages (and hours post-hatching for fed animals or post-transfer to food for starved animals at which imaging took place) shown on *x*-axis. Centre indicates median, error bars indicate quartiles (**a**, **e**). *P* values calculated by two-sided Wilcoxon rank-sum test (**a**–**e**). *n* = number of animals (shown below data points for fed, above for L1-starved) (**a**, **e**). **b**, Expression of a *tph-1* transcriptional fosmid in NSM in fed L1 animals, starved L1 animals or L1 animals fed in the presence of 20 mg ml^{-1} exogenous octopamine. Each grey dot represents averaged expression level in one animal. Magenta bar indicates median, black box represents quartiles (**b**–**d**). *n* = number of animals (shown in each column) (**b**–**d**). **c**, Expression of a *tph-1* transcriptional fosmid is not affected in ADF or NSM neurons (NSM data not shown) by exogenous tyramine in fed L1 hermaphrodites and males. TA, tyramine. **d**, *tph-1* transcript levels quantified by smFISH. Maximum intensity projection images of one half of animal to show one NSM and one ADF neuron. Merge, overlay of *tph-1* smFISH puncta onto DAPI. Number of *tph-1* smFISH puncta was normalized to number of *ric-4*/SNAP-25 synaptic protein smFISH puncta in the same neuron to control for staining fluctuations, each dot (*n* =) one neuron, shown in each column. **e**, Time-course of *tph-1* transcriptional levels in fed (solid lines) and L1-starved (dashed lines) animals. Larval stages (and hours post-hatching for fed animals or post-transfer to food for starved animals at which imaging took place) shown on *x*-axis. Asterisks (fed L3 animals) indicate that animals were imaged at different laser settings (60% of all other time points) to prevent pixel oversaturation in images: thus, we under-estimate the magnitude of the L3 serotonin spike here.

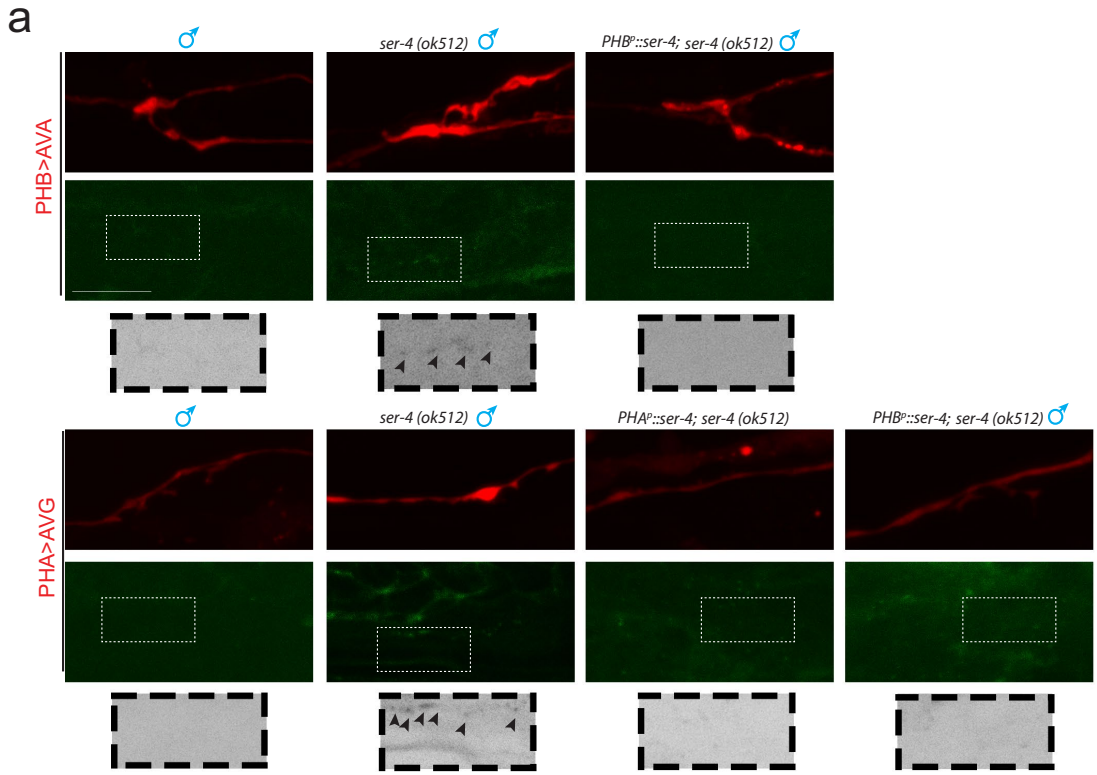


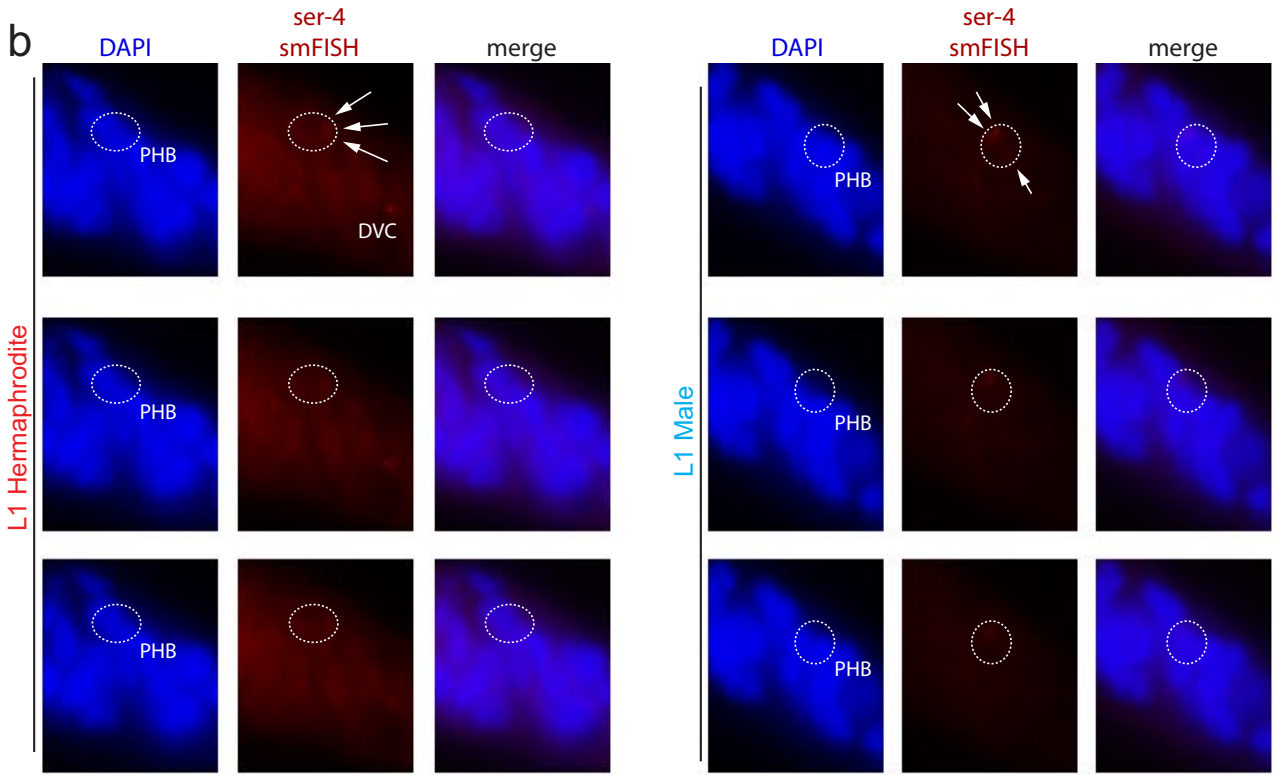
Extended Data Fig. 5 | Serotonin is downregulated by starvation and functions downstream, but not upstream, of *tbh-1* transcription. **a**, *tbh-1* is not required for the initial downregulation of *tph-1* transcription in ADF upon L1 starvation, but is required for the persistence of this downregulation into the L3 stage. Neither initial *tph-1* downregulation nor a starvation memory were apparent in NSM in a *tbh-1* null mutant. In well-fed conditions, there is no significant difference between control and *tbh-1* mutant animals. Centre indicates median, error bars indicate quartiles. Solid lines indicate continuously fed animals, dashed lines indicate L1-starved animals. *n* = number of animals, shown in each column (**a**–**c**). *P* values calculated by two-sided Wilcoxon rank-sum test (**a**–**c**). **b**, The *ser-6* octopamine receptor is required during sexual maturation to maintain *tph-1* transcription levels in ADF but not NSM under well-fed conditions, and upon starvation *tph-1* transcription levels in the *ser-6* mutant do not further decrease, supporting the necessity of *ser-6* for proper ADF starvation response. Magenta bar indicates median, black box represents quartiles (**b**, **c**). **c**, Upregulation of *tbh-1* transcription upon L1 starvation is unaffected by the addition of exogenous serotonin (5 mM), suggesting that serotonin does not act upstream of *tbh-1* upregulation (and subsequent octopamine production).



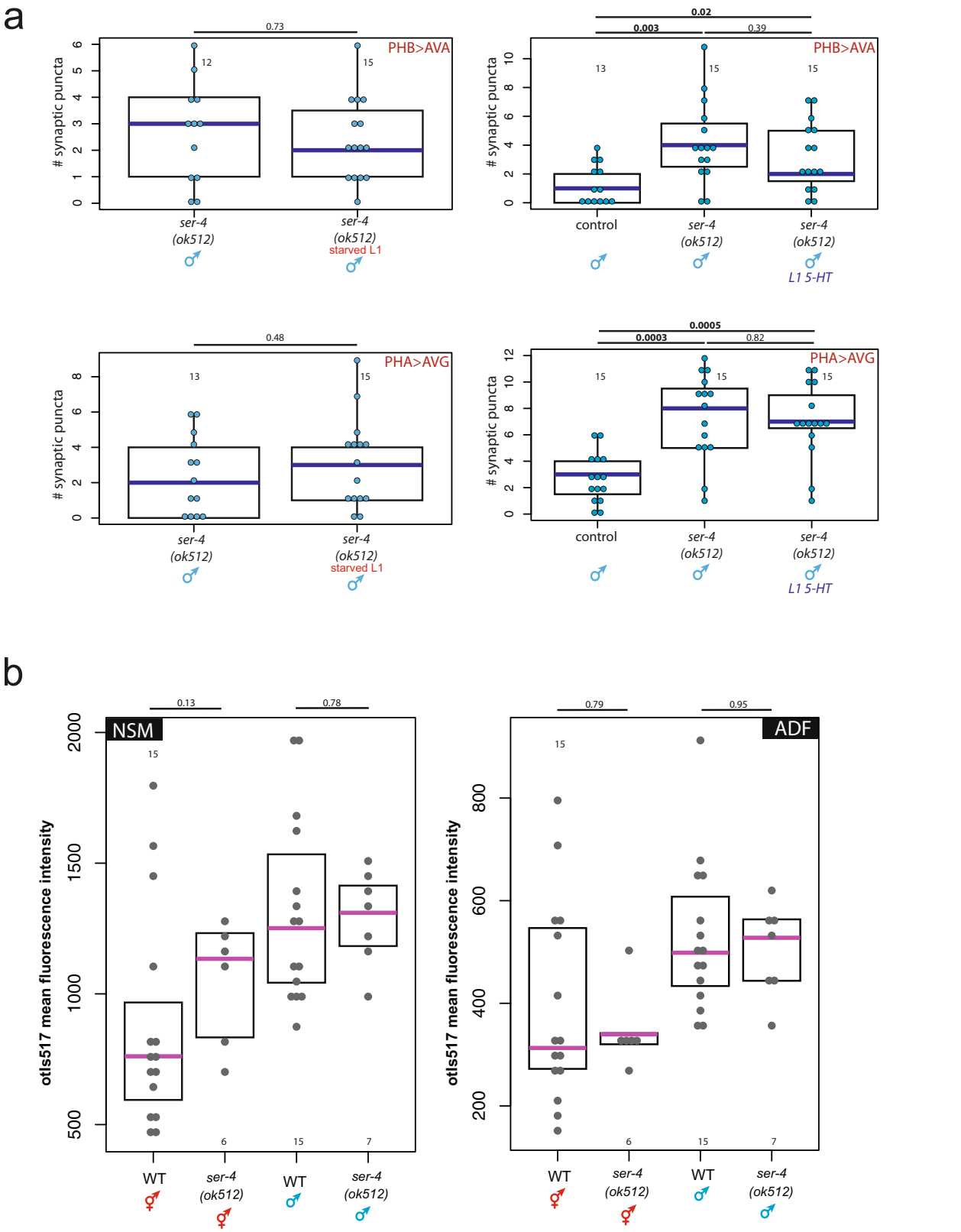
Extended Data Fig. 6 | See next page for caption.

Extended Data Fig. 6 | Regulation and neuronal migration effects of serotonin signalling. a, Effects of L1 starvation on male-specific synaptic pruning can also be rescued by exogenous serotonin during L3 (while animals are feeding). Each dot represents one adult male animal (n = number of animals, shown in each column), blue bar represents median, black box represents quartiles, vertical black bars represent range (**a**, **b**, **e**). *P* values calculated by two-sided Wilcoxon rank-sum test (**a**, **b**, **e**). **b**, *tph-1* overexpression in NSM does not rescue starvation effects on pruning. Quantification of PHB > AVA and PHA > AVG synaptic connectivity in L1-starved adult males overexpressing *tph-1* in NSM. Two independent transgenic lines were tested for each experiment; L1-starved animals without transgenic lines are siblings of transgenic animals; controls are non-starved adult males with transgenic arrays. None of the transgenic lines resulted in partial or complete rescue. **c**, Overexpression of *tph-1* under an ADF-specific promoter during L1 starvation rescues the male-specific pruning of the PHB > AVA and PHA > AVG synaptic connections. Representative images shown; for quantification and replication, see Fig. 3e and Methods. **d**, *tph-1 (n4622)* and *ttx-3 (ot22); unc-86 (n846)* mutants (in which the NSM neuron does not express *tph-1* or produce serotonin[29]) have cell body displacement, dendrite, and axon fasciculation defects in the phasmids. Overexpression of *tph-1* under an NSM-specific promoter in the *tph-1 (n4622)* mutant background rescues the severity and penetrance of these defects. Representative images of defects in the PHB neuron are shown here as inverted black and white fluorescence images. Asterisk indicates dendrite defect, arrow shows anteriorly shifted cell body, arrowhead shows fasciculation defect. Scale bars, 10 μm. Per cent of animals with visible defects categorized and quantified to the right, n = number of animals, shown in each column. *P* values calculated by Freeman–Halton extension of one-sided Fisher exact test. **e**, Overexpression of *tph-1* under an NSM-specific promoter in the *tph-1 (n4622)* mutant background (essentially, an ADF-specific *tph-1* null) results in male-specific PHB > AVA pruning defects. Of two independent NSM::*tph-1* transgenic lines, one resulted in a slight but insignificant defect in pruning, and one resulted in a substantial defect in pruning.





Extended Data Fig. 7 | *ser-4* expression in PHB is permissive for synaptic pruning. a, The normally hermaphrodite-specific PHB > AVA and PHA > AVG synaptic connections fail to prune in *ser-4* mutant males, but pruning can be rescued by cell-specific expression of *ser-4* cDNA in PHB or PHA. Top, red cytoplasmic axon label; middle, GRASP (PHB > AVA) or iBLINC signal (PHA > AVG); bottom, magnified inset of colour-inverted synaptic puncta with arrowheads to indicate puncta. Scale bars, 10 μm, all panels. Representative images shown; for quantification and replication, see Fig. 4c and Methods. **b**, *ser-4* smFISH puncta are present in PHB in both sexes at L1. Three consecutive individual *z*-slices taken from the maximum intensity projections in Fig. 4b are shown, moving laterally through each animal from top to bottom rows. Dotted circles outline the PHB nuclei (DAPI) identified using *osm-6::gfp* (see Fig. 4b). Arrowheads in the top row indicate the locations of *ser-4* puncta, which fade out of focus as the slices progress laterally. For quantification and replication, see Fig. 4b and Methods.



Extended Data Fig. 8 | *ser-4* is necessary for PHB > AVA and PHA > AVG synaptic pruning and acts downstream of the ADF serotonin signal. a, Starvation does not enhance the male synaptic pruning defect in *ser-4* mutants, and the *ser-4* mutant phenotype cannot be rescued by exogenous serotonin. Each dot represents one animal (n = number of animals, shown in each column), blue bar represents median, black box represents quartiles, vertical black bars represent range. P values calculated by two-sided Wilcoxon rank-sum test (**a**, **b**). **b**, Expression of a *tph-1* transcriptional fosmid in NSM and ADF does not significantly differ between wild-type and *ser-4* mutant L1 animals. Each dot represents the expression level of one animal, n = number of animals, shown in each column. Magenta bar indicates median, black box represents quartiles.

LETTER

https://doi.org/10.1038/s41586-018-0433-3

Alpha-kinase 1 is a cytosolic innate immune receptor for bacterial ADP-heptose

Ping Zhou[1,8], Yang She[1,2,3,8], Na Dong[4], Peng Li[1], Huabin He[1], Alessio Borio[5], Qingcui Wu[1], Shan Lu[1], Xiaojun Ding[6], Yong Cao[1], Yue Xu[1], Wenqing Gao[1], Mengqiu Dong[1], Jingjin Ding[1,2], Da-Cheng Wang[2], Alla Zamyatina[5] & Feng Shao[1,2,7]*

Immune recognition of pathogen-associated molecular patterns (PAMPs) by pattern recognition receptors often activates proinflammatory NF-κB signalling[1]. Recent studies indicate that the bacterial metabolite D-glycero-β-D-manno-heptose 1,7-bisphosphate (HBP) can activate NF-κB signalling in host cytosol[2–4], but it is unclear whether HBP is a genuine PAMP and the cognate pattern recognition receptor has not been identified. Here we combined a transposon screen in *Yersinia pseudotuberculosis* with biochemical analyses and identified ADP-β-D-manno-heptose (ADP-Hep), which mediates type III secretion system-dependent NF-κB activation and cytokine expression. ADP-Hep, but not other heptose metabolites, could enter host cytosol to activate NF-κB. A CRISPR–Cas9 screen showed that activation of NF-κB by ADP-Hep involves an ALPK1 (alpha-kinase 1)–TIFA (TRAF-interacting protein with forkhead-associated domain) axis. ADP-Hep directly binds the N-terminal domain of ALPK1, stimulating its kinase domain to phosphorylate and activate TIFA. The crystal structure of the N-terminal domain of ALPK1 and ADP-Hep in complex revealed the atomic mechanism of this ligand–receptor recognition process. HBP was transformed by host adenylyltransferases into ADP-heptose 7-P, which could activate ALPK1 to a lesser extent than ADP-Hep. ADP-Hep (but not HBP) alone or during bacterial infection induced *Alpk1*-dependent inflammation in mice. Our findings identify ALPK1 and ADP-Hep as a pattern recognition receptor and an effective immunomodulator, respectively.

Gram-negative bacteria such as *Yersinia*[5], *Salmonella*[6], *Burkholderia*[7] and enteropathogenic *Escherichia coli*[8] induce NF-κB-mediated cytokine expression in a type III secretion system (T3SS)-dependent manner. Consistently, infection of 293T cells with *Y. pseudotuberculosis* Δ6 (lacking the six T3SS effectors; Δ6 is omitted hereafter) robustly activated NF-κB-driven luciferase and eGFP reporters[5] (Fig. 1a, b). From 21,000 *Y. pseudotuberculosis* transposon mutants, we identified 37 defective in activating both reporters. Most mutations were in T3SS-encoding genes. One mutant that was more impaired than the T3SS-deficient Δ*yopB* strain had a functional T3SS with its transposon inserted in *hldE* (Fig. 1a–c and Extended Data Fig. 1a). Expression of HldE in the transposon or the Δ*hldE* mutant restored infection-induced NF-κB activation. HldE, together with GmhA and GmhB, synthesizes ADP-D-glycero-β-D-manno-heptose (ADP-DD-Hep; β is omitted hereafter) from D-sedoheptulose 7-phosphate (S7P) through D-glycero-β-D-manno-heptose 7-phosphate (H7P), HBP and D-glycero-β-D-manno-heptose 1-phosphate (H1P) (Extended Data Fig. 1b). ADP-DD-Hep and ADP-LD-Hep (ADP-L-glycero-β-D-manno-heptose) undergo interconversion, catalysed by HldD. Deletion of *gmhB* or *hldD* did not affect *Y. pseudotuberculosis*-induced activation of NF-κB, whereas deletion of *gmhA* phenocopied the Δ*hldE* strain (Fig. 1b). This seems to suggest that HBP, but not H1P or ADP-Hep, determines *Y. pseudotuberculosis*-dependent activation of NF-κB, echoing the analyses in *Neisseria meningitidis*[2]. However, *Y. pseudotuberculosis* Δ*gmhB*, unlike the Δ*hldE* and Δ*gmhA* strains, still supported ADP-Hep-dependent autotransporter heptosylation[9,10] (Extended Data Fig. 1c), because of an unknown redundancy to *gmhB* in ADP-Hep biosynthesis[11]. When electroporated into 293T cells, synthetic HBP and ADP-DD-Hep or ADP-LD-Hep—but not S7P—stimulated NF-κB activation, with ADP-Hep being the most potent (Fig. 1d). H1P was even less active than HBP[2] (Extended Data Fig. 1d). When added directly to 293T cells, only ADP-DD-Hep and ADP-LD-Hep induced activation of NF-κB and production of interleukin (IL)-8 (Fig. 1d, e). This explains the use of transfection in recent reports on HBP[2,4]. Thus, ADP-Hep is a potent and versatile PAMP.

Activation of NF-κB–eGFP reporter by extracellular ADP-Hep enabled us to carry out a fluorescence-activated cell sorting (FACS)-based genome-wide CRISPR–Cas9 screen (Extended Data Fig. 2a). Following a counterscreen against TNF stimulation, we identified *ALPK1*, *TIFA* and *TRAF6* (each hit by more than one guide RNA (gRNA)) that were required for ADP-Hep-induced NF-κB–eGFP expression (Fig. 2a and Supplementary Table 1). Upon phosphorylation at T9, TIFA forms foci to activate TRAF6-dependent NF-κB signalling[12–14]. During the preparation of this manuscript, ALPK1 was identified as contributing to activation of NF-κB in *Shigella flexneri* and *Helicobacter pylori*[3,4]. Deletion of *ALPK1* or *TIFA* (Supplementary Table 2) abolished ADP-LD-Hep-induced activation of NF-κB and expression of IL-8 (Fig. 2b, c and Extended Data Fig. 2b, c). Defective NF-κB activation in *ALPK1*$^{-/-}$ cells was restored by wild-type ALPK1 but not by its kinase-inactive K1067M mutant (Fig. 2b); *TIFA*$^{-/-}$ was rescued by wild-type TIFA but not by a T9A mutant (Extended Data Fig. 2c). ADP-LD-Hep induced phosphorylation of TIFA at T9 and formation of eGFP–TIFA foci dependent upon ALPK1 kinase activity (Fig. 2d and Extended Data Fig. 2d). ADP-LD-Hep did not affect the cytoplasmic localization of ALPK1 and induced no myosin phosphorylation[15] (Extended Data Fig. 2e, f). *ALPK1* and *TIFA* were also required for the activation of NF-κB by electroporation of ADP-Hep (Extended Data Fig. 2g).

ADP-LD-Hep triggered co-immunoprecipitation of TIFA with ALPK1 and TRAF6 (Extended Data Fig. 2h). Deletion of *ALPK1* did not affect activation of NF-κB by TNF, NOD1 or NOD2 (mediates *Salmonella*-induced NF-κB activation[16,17]) or MYD88 overexpression (Extended Data Fig. 1e). Cells lacking *NOD1* and *NOD2* showed intact NF-κB responses and TIFA foci following treatment with ADP-LD-Hep (Extended Data Fig. 1f, g). Thus, ADP-Hep activates NF-κB specifically through the ALPK1–TIFA–TRAF6 axis.

Induction of IL-8 expression by *Y. pseudotuberculosis*, which required *hldE* and *yopB*, was blocked by deletion of *ALPK1* (Fig. 2e). Infection-induced activation of NF-κB and formation of eGFP–TIFA foci required ALPK1-dependent phosphorylation of TIFA at T9 (Extended Data Fig. 3a–e). Other bacteria, such as diffuse-adhering *E. coli* (DAEC), enterotoxigenic *E. coli* (ETEC) and *Burkholderia*

[1]National Institute of Biological Sciences, Beijing, China. [2]National Laboratory of Biomacromolecules, Institute of Biophysics, Chinese Academy of Sciences, Beijing, China. [3]College of Life Sciences, University of Chinese Academy of Sciences, Beijing, China. [4]College of Animal Science and Technology, China Agricultural University, Beijing, China. [5]Department of Chemistry, University of Natural Resources and Life Sciences, Vienna, Austria. [6]Beijing Mingde Zhengkang Technologies Co., Ltd., Beijing, China. [7]Tsinghua Institute of Multidisciplinary Biomedical Research, Tsinghua University, Beijing, China. [8]These authors contributed equally: Ping Zhou, Yang She. *e-mail: shaofeng@nibs.ac.cn

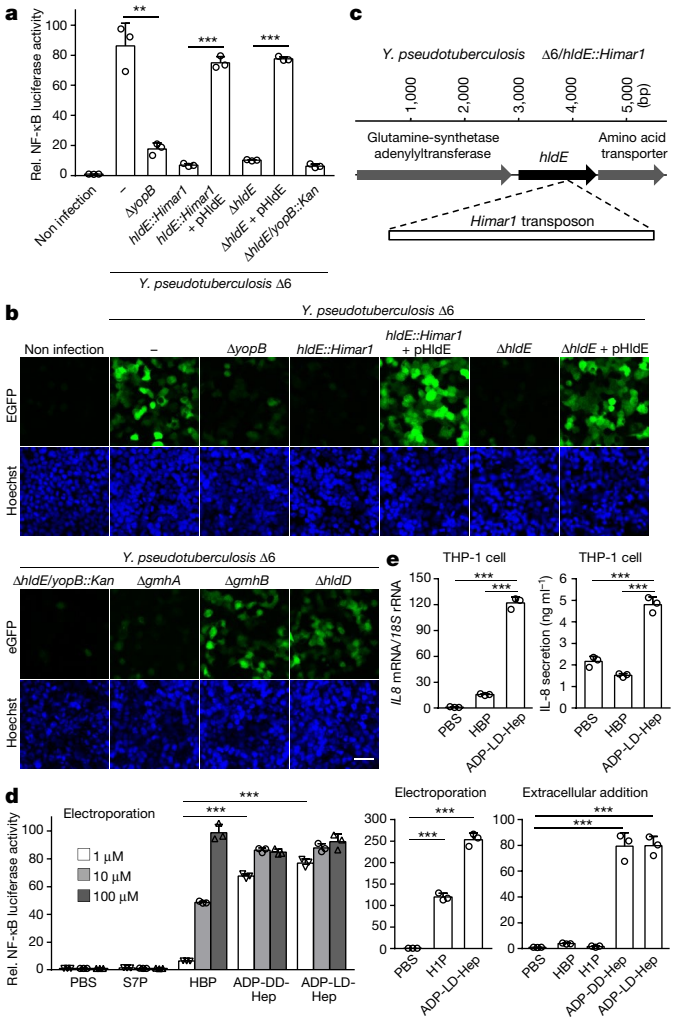


Fig. 1 | Transposon screen of *Yersinia* T3SS-dependent NF-κB activation identifies ADP-Hep as a PAMP. a, **b**, 293T cells were infected with indicated *Yersinia* strains. pHldE is a plasmid expressing HldE. **c**, Transposon insertion in *hldE*. **d**, **e**, 293T (**d**) or PMA-differentiated THP-1 cells (**e**) were electroporated (**d**) or extracellularly treated (**d**, **e**) with indicated sugars. *IL8* mRNA (also known as *CXCL8*, shown relative to *18S* rRNA) and IL-8 secretion were measured by quantitative real-time PCR (qPCR) and enzyme-linked immunosorbent assay (ELISA), respectively. NF-κB activation was measured by luciferase activity (**a**, **d**) or eGFP reporter expression (**b**) (scale bar, 50 μm). Data shown as mean ± s.d. from three technical replicates (**a**, **d**, **e**), and representative of three (**a**, **b**, **d**) and two (**e**) independent experiments. **a**, **d**, **e**, Two-tailed unpaired Student's *t*-test (**$P < 0.01$, ***$P < 0.001$).

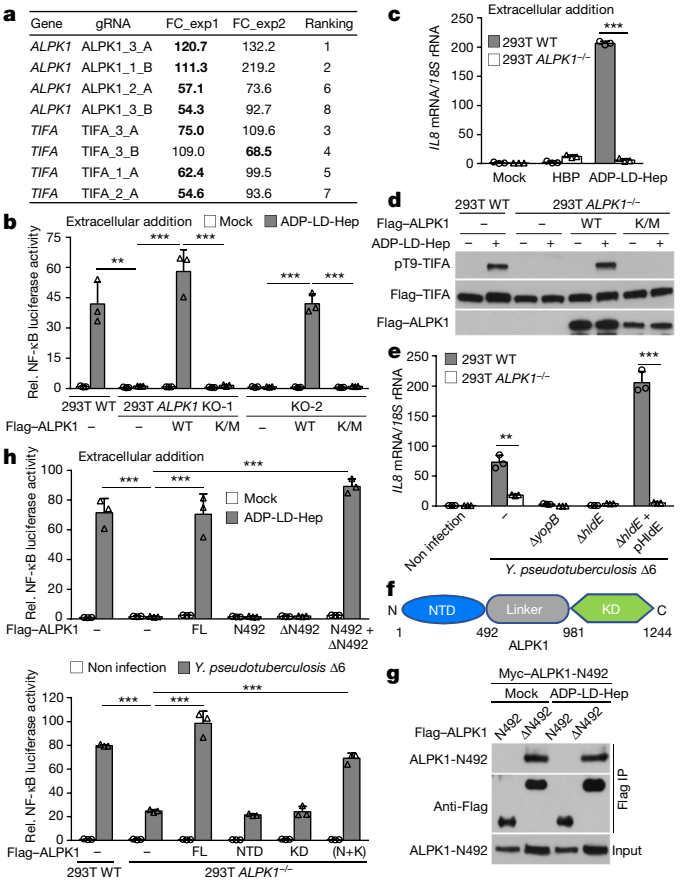


Fig. 2 | CRISPR–Cas9 screens identify an ALPK1–TIFA axis that mediates activation of NF-κB induced by ADP-Hep or *Yersinia*. a, Top eight gRNA hits from CRISPR–Cas9 screen of ADP-LD-Hep-induced NF-κB activation. **b**–**e**, **h**, Wild-type or *ALPK1*$^{-/-}$ 293T cells expressing the indicated ALPK1 mutants were treated with ADP-LD-Hep (**b**–**d**, **h**) or HBP (**c**), or infected with *Y. pseudotuberculosis* (**e**, **h**). **d**, Anti-Flag and anti-pT9-TIFA immunoblots of 293T cells. **f**, Domain organization of ALPK1. **g**, Co-immunoprecipitation of ALPK1 N- and C-terminal regions from 293T cells treated with or without ADP-LD-Hep. NF-κB activation and *IL8* mRNA were assessed by luciferase reporter activity (**b**, **h**) or qRT–PCR (**c**, **e**), respectively. KO-1/2, two *ALPK1*$^{-/-}$ clones. Data shown as mean ± s.d. from three technical replicates (**b**, **c**, **e**, **h**) and representative of three (**a**, **b**, **d**, **h**, **g**) or two (**c**, **e**) independent experiments. **b**, **c**, **e**, **h**, Two-tailed unpaired Student's *t*-test (**$P < 0.01$, ***$P < 0.001$).

cenocepacia also triggered NF-κB activation mediated by the ALPK1–TIFA axis in an *hldE*-dependent manner (Extended Data Fig. 3f, g). Thus, sensing of ADP-Hep by ALPK1 is not limited to T3SS-containing bacteria.

ALPK1 contains an α-helical domain and a kinase domain linked by an extensive unstructured region (Fig. 2f). ALPK1-N492 (residues 1–492) and ALPK1-ΔN492 (lacking these residues) were co-immunoprecipitated, independently of ADP-LD-Hep (Fig. 2g). Co-expression of ALPK1-N492 and ALPK1-ΔN492, or minimally the N-terminal and kinase domains of ALPK1 (ALPK1-NTD (1–473) and ALPK1-KD (959–1244), respectively), was sufficient to allow ADP-LD-Hep or *Y. pseudotuberculosis* to induce activation of NF-κB and phosphorylation of TIFA (Fig. 2h and Extended Data Fig. 4a–e). Unexpectedly, a purified complex of ALPK1-NTD and ALPK1-KD (ALPK1-(N+K); Extended Data Fig. 5a) could directly phosphorylate T9 of TIFA (Fig. 3a). By contrast, activation of ALPK1-(N+K) in mammalian cells required infection or stimulation by ADP-Hep.

Notably, ALPK1-(N+K) from *E. coli* Δ*hldE* did not phosphorylate TIFA (Fig. 3a). Small-molecule extracts from wild-type but not Δ*hldE* mutant *E. coli*-derived ALPK1-NTD, when added to 293T cells, potently stimulated ALPK1-dependent phosphorylation of TIFA and activation of NF-κB, the latter of which correlated with the amount of ALPK1-NTD (Fig. 3b, c).

High-performance liquid chromatography (HPLC) of the small-molecule extracts identified one active fraction (no. 6, the only one with high ultraviolet absorption) (Extended Data Fig. 5b–d). Mass spectrometry of fraction 6 uncovered three dominant ions with mass-to-charge ratios (*m*/*z*) of 619.8, 347.9 and 427.8, matching those of ADP-Hep, AMP and ADP, respectively (Extended Data Fig. 5e). The presumed ADP-Hep ion showed a similar retention time and fragmentation pattern to synthetic ADP-Hep (Extended Data Fig. 5f, g). The mass of native ALPK1-NTD exceeded that of denatured ALPK1-NTD by 619.32 Da (one ADP-Hep) (Fig. 3d). Direct binding of *E. coli* Δ*hldE*-derived apo-ALPK1-(N+K) to ADP-LD-Hep (but not S7P) was readily detected (Extended Data Fig. 5h).

We determined the 2.59 Å crystal structure of ALPK1-NTD purified from wild-type *E. coli* (Extended Data Table 1a). Each asymmetric

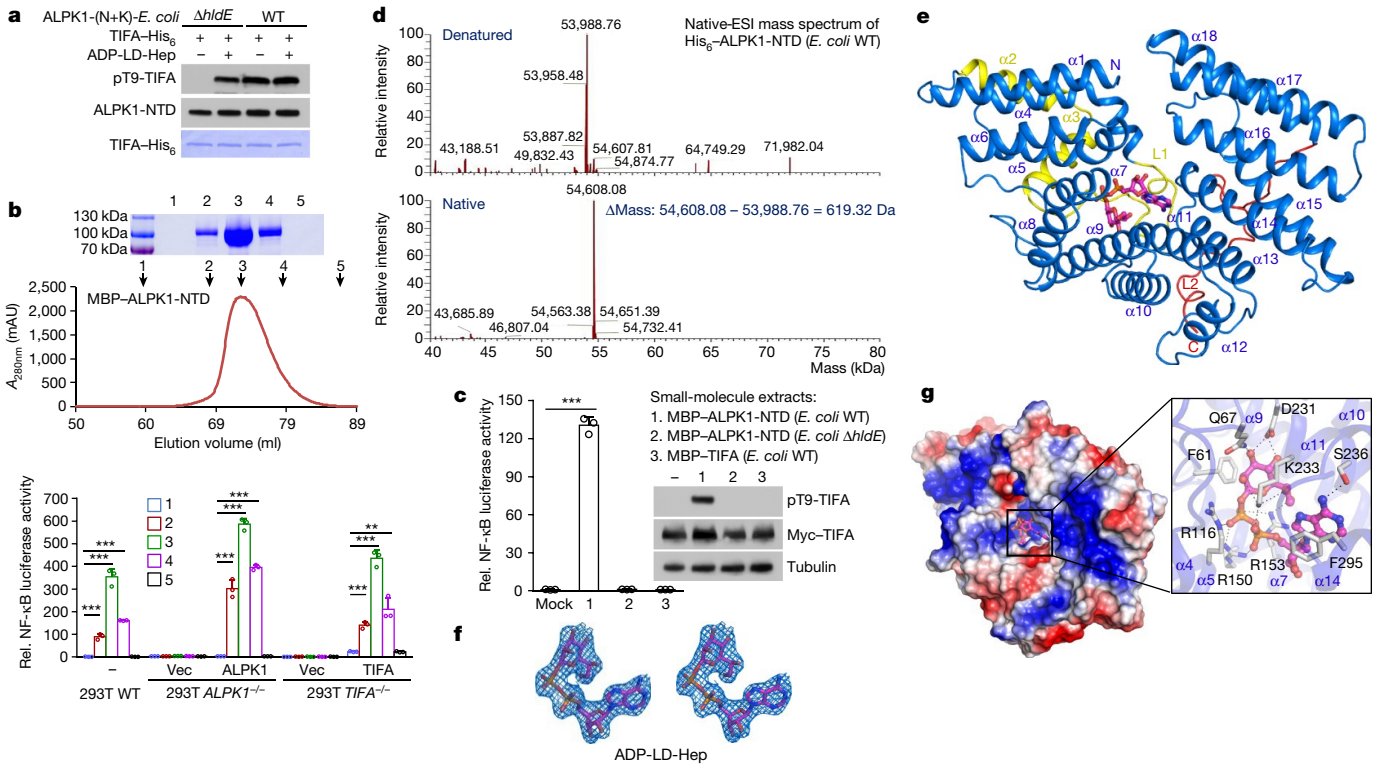


Fig. 3 | ADP-Hep binding to ALPK1-NTD and crystal structure of the binding complex. a, ALPK1-(N+K) purified from indicated *E. coli* strains was incubated with histidine-tagged TIFA (TIFA–His_6) in the presence or absence of ADP-LD-Hep. **b**, Gel-filtration chromatography of MBP–ALPK1-NTD purified from wild-type *E. coli*. **b**, **c**, 293T cells were treated with small-molecule extracts from fractions 1–5 (**b**) or from TIFA or MBP–ALPK1-NTD (**c**). Activation of NF-κB was assessed by luciferase reporter assay (mean ± s.d. from three technical replicates); two-tailed unpaired Student's *t*-test, $***P < 0.001$. **d**, Electrospray ionization mass spectrometry of native and denatured ALPK1-NTD purified from wild-type *E. coli*. **e**, Structure of the ALPK1-NTD–ADP-Hep complex. **f**, Stereo diagram of the simulated annealing F_o-F_c omit map of ADP-LD-Hep contoured at 3σ. **g**, Surface representation of ADP-Hep-bound ALPK1-NTD and overview of the binding pocket. Black dashed lines, hydrogen bonds. **a**, **c**, Anti-pT9-TIFA immunoblotting of TIFA phosphorylation. Data are representative of three (**a**–**c**) or two (**d**) independent experiments.

unit contains nine molecules (A–I). Molecule A has the highest-quality density map and was used to build the final model, which lacks only the first methionine. The structure bears 18 helices (α1 to α18), forming seven antiparallel pairs (α1–α4, α5–α6, α7–α8, α9–α10, α11–α13, α14–α15, and α16–α17) (Fig. 3e). The α1–α4 pair features an insertion containing α2, α3 and loop L1. α18, which precedes the L2 tail, flanks the outer surface of the α16–α17 pair. The one-by-one-packed seven helix pairs form a right-hand solenoid (Fig. 3e), resembling a

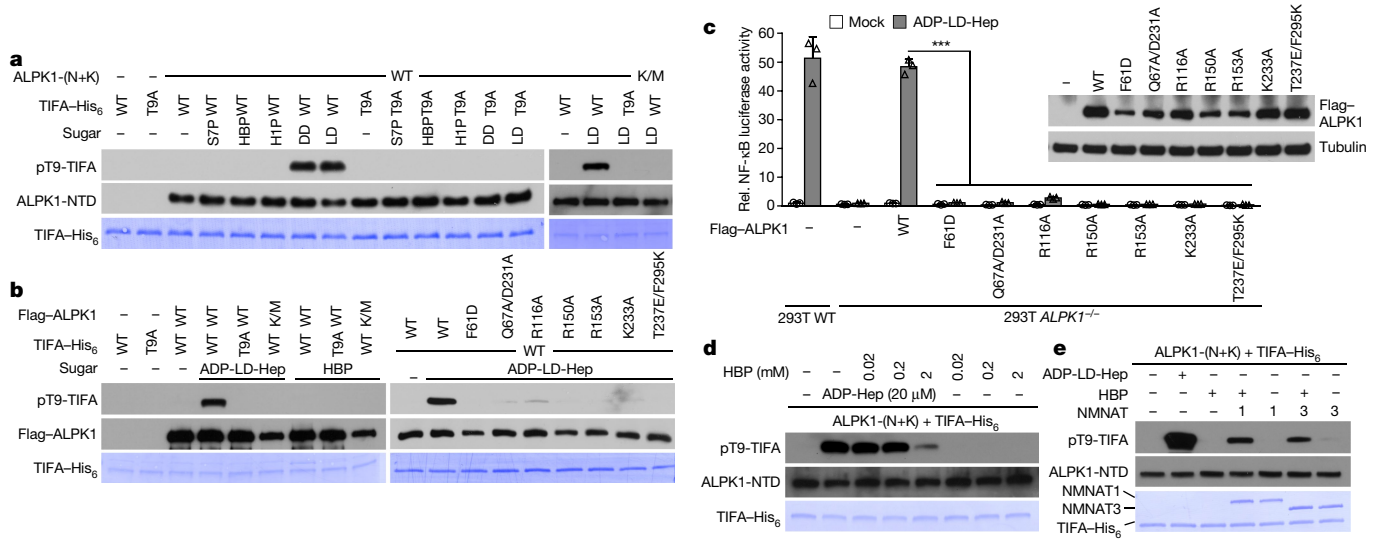


Fig. 4 | Binding of ADP-Hep but not HBP activates ALPK1 in vitro. a, **b**, **d**, **e**, Apo-ALPK1-(N+K) (**a**, **d**, **e**) or 293T cell-purified Flag–ALPK1 (wild-type or indicated mutants; K/M denotes kinase-inactive K1067M mutant) (**b**) were incubated with TIFA–His_6 in the presence of different sugars. A titrating amount of HBP was added (**d**); HBP was pre-treated with NMNAT1 or NMNAT3 (**e**). Proteins were purified from *E. coli* Δ*hldE*; TIFA phosphorylation was assessed by anti-pT9-TIFA immunoblotting. DD, ADP-DD-Hep; LD, ADP-LD-Hep. **c**, Wild-type or *ALPK1*$^{-/-}$ 293T cells expressing the indicated ALPK1 mutants were stimulated with ADP-LD-Hep. Luciferase assay of NF-κB activation is shown as mean ± s.d. from three technical replicates (two-tailed unpaired Student's *t*-test, $***P < 0.001$). All data are representative of three independent experiments.

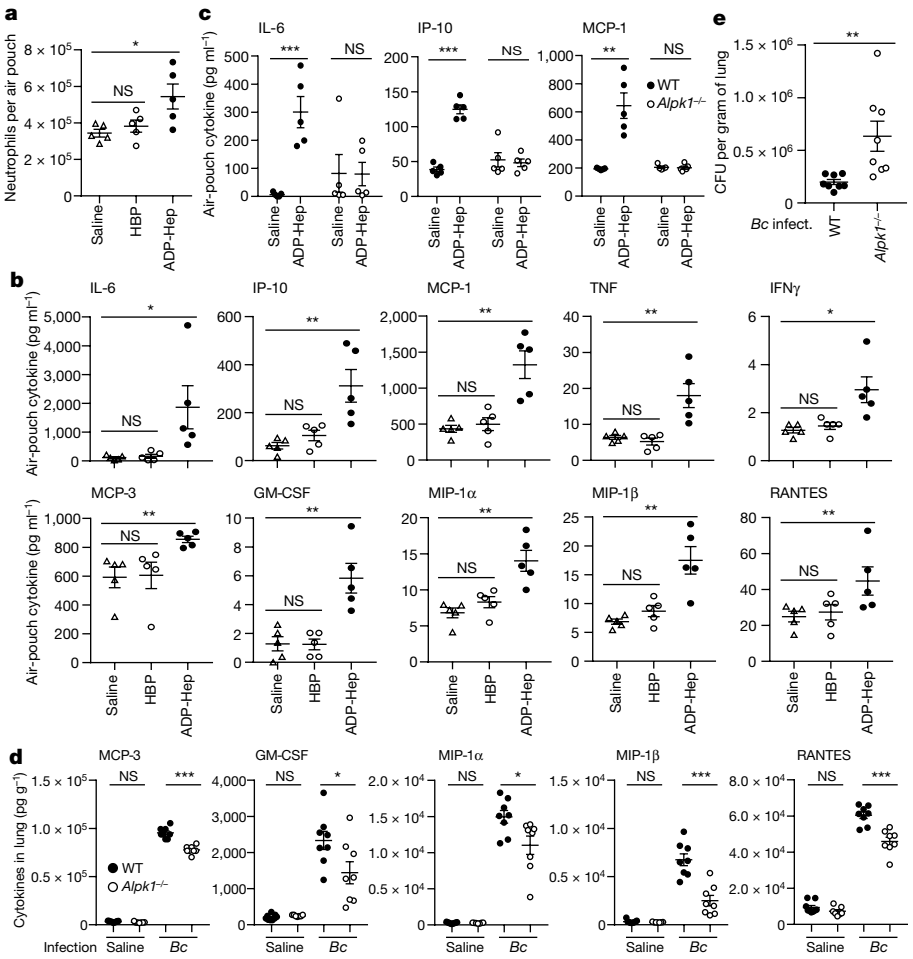


Fig. 5 | ADP-Hep and *B. cenocepacia* infection induce ALPK1-dependent inflammatory responses in mice. a–c, HBP or ADP-LD-Hep (2 mg kg^{-1}) was injected into the dorsal air pouches of wild-type (**a**–**c**) or *Alpk1*$^{-/-}$ mice (**c**). **d**, **e**, Mice were intratracheally infected for 24 h with *B. cenocepacia* (*Bc*) J2315 (5×10^7). **a**, Neutrophil counts in the air pouch. **b**–**d**, Cytokine concentrations in air pouch washes (**b**, **c**) or lung homogenates (**d**) determined by multiplex immunoassay (**b**, **d**) or ELISA (**c**). **e**, Numbers of bacteria in lungs. Data shown as mean ± s.e.m. (two-tailed unpaired Student's *t*-test, $*P < 0.05$, $**P < 0.01$, $***P < 0.001$; NS, not significant). *n* (biologically independent animals) = 5 in **a**–**c**, 9 for saline/wild-type, 7 for saline/*Alpk1*$^{-/-}$ and 8 for *B. cenocepacia* in **d**, **e**. All data are representative of two independent experiments.

tetratricopeptide repeat (TPR) domain structure. The top hits from a Dali search were all TPR domains, despite lacking sequence homology to ALPK1-NTD (Extended Data Table 1b).

The concave side of the ALPK1-NTD solenoid has a narrow pocket (Fig. 3e). The pocket has high-quality electron densities specifically for ADP-LD-Hep (although it is spacious enough for ADP-DD-Hep; Fig. 3f). The two ADP phosphates are clamped by R116, R150, R153 and K233 through a hydrogen-bond network: two bonds between R116 (or R153) and β-phosphate, two between R150 and both phosphates, and one between K233 and α-phosphate (Fig. 3g). The ADP adenosine has a stacking interaction with F295 and two hydrogen bonds with S236 and T237. The heptose C3 and C4 hydroxyls are anchored by Q67 and D231, C2 and C6 are contacted by K233, and C7 is fixed by R153. The heptose backbone has a stacking interaction with F61. These ADP-Hep-binding residues are conserved in ALPK1 of other vertebrates (Extended Data Fig. 4f).

Apo-ALPK1-(N+K) could not phosphorylate TIFA. It was activated upon incubation with ADP-LD or DD-Hep, but not S7P, HBP or H1P (Figs. 3a, 4a). The inactivity of S7P is consistent with its lack of binding to ALPK1-NTD (Extended Data Fig. 5h) and failure to induce NF-κB activation (Fig. 1d). Full-length ALPK1 purified from 293T cells also responded to ADP-LD-Hep but not to HBP (Fig. 4b). The R116A, R150A, R153A or K233A mutant forms of ALPK1, which are expected to have impaired binding to the phosphates of ADP-Hep, resisted activation by ADP-Hep (Fig. 4b). Similar results were obtained with Q67A/D231A, T237E/F295K and F61D mutations, which affect heptose or adenosine binding. These ALPK1 mutants could not mediate activation of NF-κB by ADP-LD-Hep (Fig. 4c). The requirement of AMP-contacting residues for activation of ALPK1 explains why HBP is inactive in vitro. Thus, ALPK1 is a functional receptor specifically for ADP-Hep.

HBP could bind apo-ALPK1-(N+K) but with a lower affinity than ADP-Hep (Extended Data Fig. 6a); a 100-fold excess of HBP was needed for it to compete with ADP-LD-Hep to activate ALPK1 (Fig. 4d and Extended Data Fig. 6b). ALPK1-(N+K) incubated with ADP-LD-Hep could bind and phosphorylate a GST-fused N-terminal 15-residue peptide of TIFA and catalyse ATP hydrolysis (Extended Data Fig. 6c–e). By contrast, ALPK1-(N+K) incubated with HBP showed neither substrate recognition nor catalysis of ATP hydrolysis. Chemical cross-linking coupled with mass spectrometry identified ten peptide pairs between the NTD and KD molecules of apo-ALPK1-(N+K) (Extended Data Fig. 6f and Supplementary Table 3). ALPK1-(N+K) incubated with HBP showed a similar pattern, but ALPK1-(N+K) incubated with ADP-LD-Hep had only six crosslinks. The four connections to K1149 of ALPK1-KD and one to the nearby K1140, which were partially lost in HBP-incubated ALPK1-(N+K), all disappeared in ADP-LD-Hep-incubated ALPK1-(N+K). The two lysines are predicted to be near the kinase catalytic cleft[18] (Extended Data Fig. 6g), which participates in substrate binding. Thus, binding of ADP-LD-Hep to ALPK1-NTD renders the catalytic cleft more

exposed, probably by inducing a larger conformational change than the nonproductive HBP binding.

Activation of NF-κB induced by HBP transfection required phosphorylation of TIFA by ALPK1[3,4] (Extended Data Fig. 7a, b). We considered how HBP activates ALPK1 in cells. After noticing a side reaction of HldE adenylyltransferase with HBP, we found that host-derived adenylyltransferases—such as nicotinamide nucleotide adenylyltransferases—could convert HBP into ADP-heptose 7-P, which was competent (albeit less than ADP-Hep) to activate ALPK1 and the downstream NF-κB response (Fig. 4e and Extended Data Fig. 7c–k; see Supplementary Text). We also found that ALPK1(Q67A), ALPK1(Y68A) or the double mutant lost the response to HBP electroporation but remained competent to mediate ADP-LD-Hep-induced activation of NF-κB (Extended Data Fig. 8a, b). Cells expressing these mutants also showed defective responses to ADP-heptose 7-P (Extended Data Fig. 8c). Purified ALPK1(Q67A/Y68A) was activated by ADP-LD-Hep but not ADP-heptose 7-P or HBP (Extended Data Fig. 8d). Small-molecule extracts from cells electroporated with HBP could activate wild-type ALPK1 but not ALPK1(Q67A/Y68A) (Extended Data Fig. 8e, f). Thus, cytosolic HBP might be metabolized into ADP-heptose 7-P to activate ALPK1. Moreover, 293T cells expressing ALPK1(Q67A/Y68A) showed a normal NF-κB response to *Y. pseudotuberculosis* (Extended Data Fig. 8g), confirming the detection of ADP-Hep by the host during infection.

Injection of ADP-LD-Hep, but not of HBP, into the dorsal air pouches of mice induced massive neutrophil recruitment (Fig. 5a). Several NF-κB-controlled cytokines and chemokines, including IL-6, TNF, IP-10, MCP-1, MCP-3, IFNγ, GM-CSF, MIP-1α, MIP-1β and RANTES, were highly elevated in the air pouches (Fig. 5b and Extended Data Fig. 9a). ADP-LD-Hep also increased the serum levels of GRO-α, IP-10 and MCP-1 (Extended Data Fig. 9b, c). HBP injection affected neither local nor systemic productions of these inflammatory mediators (Fig. 5b and Extended Data Fig. 9a–c). *Alpk1*$^{-/-}$ mice injected with ADP-LD-Hep showed no increase in cytokine production (Fig. 5c and Extended Data Fig. 9d). Mice were also infected with *B. cenocepacia*, which is known to trigger lung inflammation. Consistent with the cell culture data (Extended Data Fig. 3f, g), infection with *B. cenocepacia* increased the expression of MCP-3, GM-CSF, MIP-1α and β, and RANTES in the lungs of wild-type mice, and these responses were compromised in *Alpk1*$^{-/-}$ mice (Fig. 5d). *Alpk1*$^{-/-}$ mice showed a higher bacterial load in the lungs than wild-type mice (Fig. 5e). These data emphasize the functional relevance of recognition of ADP-Hep by ALPK1.

We have shown that ADP-Hep is permeable to mammalian cells and can be exploited as an immunomodulator or a vaccine adjuvant. The pattern recognition receptor function of ALPK1 highlights the versatile mechanisms underlying cytosolic detection of bacteria[19] and will stimulate research into other alpha kinases[15]. ADP-Hep is present in all Gram-negative and some Gram-positive bacteria[20]; ALPK1 is also widely expressed. Recognition of ADP-Hep by ALPK1, as with recognition of LPS by TLR4 and caspase-11, represents a generic form of innate sensing, which mediates immune responses to diverse bacterial pathogens. These include *N. meningitidis*[2], DAEC and ETEC, which are extracellular and possess no injection systems, as well as *Yersinia* spp., *B. cenocepacia*, *S. flexneri*[3] and *H. pylori*[4,21,22].

Online content

Any Methods, including any statements of data availability and Nature Research reporting summaries, along with any additional references and Source Data files, are available in the online version of the paper at https://doi.org/10.1038/s41586-018-0433-3

Received: 31 December 2017; Accepted: 23 July 2018;
Published online 15 August 2018.

1. Pandey, S., Kawai, T. & Akira, S. Microbial sensing by Toll-like receptors and intracellular nucleic acid sensors. *Cold Spring Harb. Perspect. Biol.* **7**, a016246 (2014).
2. Gaudet, R. G. et al. Cytosolic detection of the bacterial metabolite HBP activates TIFA-dependent innate immunity. *Science* **348**, 1251–1255 (2015).
3. Milivojevic, M. et al. ALPK1 controls TIFA/TRAF6-dependent innate immunity against heptose-1,7-bisphosphate of gram-negative bacteria. *PLoS Pathog.* **13**, e1006224 (2017).
4. Zimmermann, S. et al. ALPK1- and TIFA-dependent innate immune response triggered by the *Helicobacter pylori* type IV secretion system. *Cell Rep.* **20**, 2384–2395 (2017).
5. Auerbuch, V., Golenbock, D. T. & Isberg, R. R. Innate immune recognition of *Yersinia pseudotuberculosis* type III secretion. *PLoS Pathog.* **5**, e1000686 (2009).
6. Bruno, V. M. et al. *Salmonella typhimurium* type III secretion effectors stimulate innate immune responses in cultured epithelial cells. *PLoS Pathog.* **5**, e1000538 (2009).
7. Hii, C. S. et al. Interleukin-8 induction by *Burkholderia pseudomallei* can occur without Toll-like receptor signaling but requires a functional type III secretion system. *J. Infect. Dis.* **197**, 1537–1547 (2008).
8. Litvak, Y. et al. Epithelial cells detect functional type III secretion system of enteropathogenic *Escherichia coli* through a novel NF-κB signaling pathway. *PLoS Pathog.* **13**, e1006472 (2017).
9. Lu, Q. et al. An iron-containing dodecameric heptosyltransferase family modifies bacterial autotransporters in pathogenesis. *Cell Host Microbe* **16**, 351–363 (2014).
10. Yao, Q. et al. A structural mechanism for bacterial autotransporter glycosylation by a dodecameric heptosyltransferase family. *eLife* **3**, e03714 (2014).
11. Kneidinger, B. et al. Biosynthesis pathway of ADP-L-*glycero*-β-D-*manno*-heptose in *Escherichia coli*. *J. Bacteriol.* **184**, 363–369 (2002).
12. Takatsuna, H. et al. Identification of TIFA as an adapter protein that links tumor necrosis factor receptor-associated factor 6 (TRAF6) to interleukin-1 (IL-1) receptor-associated kinase-1 (IRAK-1) in IL-1 receptor signaling. *J. Biol. Chem.* **278**, 12144–12150 (2003).
13. Huang, C. C. et al. Intermolecular binding between TIFA-FHA and TIFA-pT mediates tumor necrosis factor alpha stimulation and NF-κB activation. *Mol. Cell. Biol.* **32**, 2664–2673 (2012).
14. Ea, C. K., Sun, L., Inoue, J. & Chen, Z. J. TIFA activates IκB kinase (IKK) by promoting oligomerization and ubiquitination of TRAF6. *Proc. Natl Acad. Sci. USA* **101**, 15318–15323 (2004).
15. Heine, M. et al. α-kinase 1, a new component in apical protein transport. *J. Biol. Chem.* **280**, 25637–25643 (2005).
16. Keestra, A. M. et al. Manipulation of small Rho GTPases is a pathogen-induced process detected by NOD1. *Nature* **496**, 233–237 (2013).
17. Keestra, A. M. et al. A *Salmonella* virulence factor activates the NOD1/NOD2 signaling pathway. *MBio* **2**, e00266-11 (2011).
18. Yamaguchi, H., Matsushita, M., Nairn, A. C. & Kuriyan, J. Crystal structure of the atypical protein kinase domain of a TRP channel with phosphotransferase activity. *Mol. Cell* **7**, 1047–1057 (2001).
19. Zhao, Y. & Shao, F. Diverse mechanisms for inflammasome sensing of cytosolic bacteria and bacterial virulence. *Curr. Opin. Microbiol.* **29**, 37–42 (2016).
20. Tang, W. et al. D-Sedoheptulose-7-phosphate is a common precursor for the heptoses of septacidin and hygromycin B. *Proc. Natl Acad. Sci. USA* **115**, 2818–2823 (2018).
21. Gall, A., Gaudet, R. G., Gray-Owen, S. D. & Salama, N. R. TIFA signaling in gastric epithelial cells initiates the *cag* type 4 secretion system-dependent innate immune response to *Helicobacter pylori* infection. *MBio* **8**, e01168-17 (2017).
22. Stein, S. C. et al. *Helicobacter pylori* modulates host cell responses by CagT4SS-dependent translocation of an intermediate metabolite of LPS inner core heptose biosynthesis. *PLoS Pathog.* **13**, e1006514 (2017).

Acknowledgements We thank R. Zhou, X. Tan, F. Wang, H. Huang, X. Tian, Y. Ma, J. Li and the Metabolomics Facility at Tsinghua University Technology Center for Protein Sciences for technical assistance. The work was supported by grants from China NSFC (81788101), MOST of China (2017YFA0505900 and 2016YFA0501500), the Chinese Academy of Sciences (XDB08020202) and the Austrian Science Fund (P-28915).

Reviewer information *Nature* thanks S. Hur and the other anonymous reviewer(s) for their contribution to the peer review of this work.

Author contributions P.Z. and F.S. conceived the study; P.Z. did most experiments; Y.S., supervised by J.D. and D.-C.W., determined the structures and performed some biochemical assays; H.H. and W.G. performed mouse assays; S.L., X.D., Y.C. and M.D. did mass spectrometry experiments; A.B. and A.Z. synthesized the sugars; N.D., Q.W., Y.X. and P.L. provided technical assistance; P.Z., J.D. and F.S. analysed the data and wrote the manuscript. All authors discussed the results and commented on the manuscript.

Competing interests The authors declare no competing interests.

Additional information

Extended data is available for this paper at https://doi.org/10.1038/s41586-018-0433-3.
Supplementary information is available for this paper at https://doi.org/10.1038/s41586-018-0433-3.

Correspondence and requests for materials should be addressed to F.S.

METHODS

Plasmids, antibodies and reagents. DNA for *hldE* was amplified from *Y. pseudotuberculosis* IP2666 and inserted into the pBAD24 vector for rescue expression in *Yersinia*. DNAs for *gmhA*, *hldE* and *gmhB* were amplified from *E. coli* DH5α and cloned into pET28a-6×His-SUMO or pQE-80L vectors for recombinant expression in *E. coli*. cDNA for human *ALPK1* was amplified from 293T cDNA reverse transcripts. cDNAs for *TIFA* and *MYD88* were amplified from a HeLa cDNA library. cDNA for myosin I was from an ORF library from Invitrogen (clone ID: IOH29181). For expression in mammalian cells, cDNAs of indicated genes were cloned into the pCS2-3×Flag or pCS2-6×Myc vectors for transient transfection, and for stable expression the cDNAs were inserted into the FUIPW lentiviral vector with an N-terminal eGFP or mCherry tag. For recombinant expression in bacteria, indicated *ALPK1* cDNA fragments were inserted into the pMAL-c2X, pQE-80L, pGEX-6p-2, pET28a-6×His-SUMO or pACSUMO (the origin of pET28a-6×His-SUMO was replaced with p15A derived from pACYC vector) vectors, and *TIFA* cDNA was cloned into pET-22b vector with a C-terminal 6×His tag. cDNAs for human *NMNAT1* (isoform 1) and *NMNAT3* (isoform 4) were synthesized by our in-house gene synthesis facility and cloned into the pGEX-6p-2 vector for recombinant expression in bacteria. The NOD1 and NOD2 expression plasmids were kindly provided by G. Nunez (University of Michigan) and the TRAF6 expression plasmid was previously described[23]. Flag-tagged AIDA-I fragment (GST–AIDA-$I_{50–600}$–Flag) and AAH expression plasmids have also previously been described[9]. The pNF-κB-eGFP reporter plasmid (pNL2.2-BII-5RE-eGFP) used in this study was generated by inserting five copies of κB sites into the multiple cloning sites in pNL2.2 (Promega) and at the same time replacing the *Nluc* cDNA with that of eGFP. All luciferase assay plasmids were described in our previous publications[23–26]. The lentiCas9-Blast and lentiGuide-Puro plasmids used for generating the knockout cells were obtained from Addgene. All truncations and point mutations were generated by the standard PCR cloning strategy. All plasmids were verified by DNA sequencing.

The anti-AAH antibody has previously been described[9]. The rabbit polyclonal antibody for ALPK1 was from GeneTex (#GTX87015). Antibodies for tubulin (T5168) and Flag (M2) were from Sigma-Aldrich. The anti-Myc monoclonal antibody (9E10) was from Covance. The rabbit polyclonal antibody (anti-pT9-TIFA) against T9-phosphorylated human TIFA was developed by Abcam as a collaborative project using a synthetic phospho-peptide antigen, from which a monoclonal antibody was generated (ab214815).

HBP, ADP-DD-Hep and ADP-LD-Hep have previously been synthesized[27,28]. D-Sedoheptulose 7-phosphate (S7P) (#78832) was purchased from Sigma-Aldrich. Recombinant human TNF (#rcyc-htnfa), C12-iE-DAP (#tlrl-c12dap), and muramyl dipeptide (MDP) (#tlrl-mdp) were InvivoGen products. ELISA kits for human IL-8, mouse IL-6 and MCP1 were purchased from Dakewe Biotech; ELISA kits for mouse GRO-α and mouse IP-10 were from R&D Systems and Neobioscience, respectively. All other chemical reagents used were from Sigma-Aldrich unless noted.

Bacterial strains and infection. *Y. pseudotuberculosis* IP2666 Δ6 lacking six T3SS effector proteins (YopH, YopE, YopM, YopO, YopJ and YopT) was provided by R. R. Isberg (Tufts University School of Medicine). *S. flexneri* 2a strain 2457T, DAEC 2787, ETEC H10407 and *B. cenocepacia* J2315 were as described[9,29,30]. *E. coli* BL21 (DE3) strain was used for recombinant protein expression. Bacterial deletion mutants were generated by using the λ Red recombineering technology as previously described[31]. For complementation of *Y. pseudotuberculosis* IP2666 Δ6/Δ*hldE*, 0.2% L-arabinose was used to induce HldE expression. To determine the biosynthesis of ADP-Hep in *Y. pseudotuberculosis* mutants, plasmids expressing GST–AIDA-$I_{50–600}$–Flag and AAH were transformed into the bacteria, and heptosylation of AIDA-I was detected with the ECL glycoprotein detection kit (GE Healthcare) as previously described[9]. Low calcium-induced type III secretion assay was performed by following an established protocol[32].

For cell culture infection, bacteria were cultured at 30 °C (for *Y. pseudotuberculosis*) or 37 °C (for *S. flexneri*) in Luria-Bertani (LB) broth with shaking until OD_{600} reached 1.5. 293T cells were seeded in 96- or 24-well plates and cultured for 16 h before the infection (MOI 50). The infection was facilitated by centrifugation at 800*g* for 5 min at room temperature. One and a half hours after the infection, the cell culture medium was placed with refresh Dulbecco's modified Eagle's medium (DMEM) containing 34 μg ml^{-1} chloramphenicol to prevent over-proliferation of both extra- and intracellular bacteria. Infected cells were further cultured for 12 h or 4 h before being subjected to NF-κB luciferase/eGFP reporter assays or indicated immunoblotting assays, respectively.

Bacterial transposon screen. Ampicillin-resistant *Y. pseudotuberculosis* IP2666 Δ6 was obtained by transformation with the pBAD24-mCherry plasmid. The *Himar1* mariner transposon vector pSC123[33] (kindly provided by A. Rietsch, Case Western Reserve University and S. Lory, Harvard Medical School) was used to generate the mutant library of *Y. pseudotuberculosis* IP2666 Δ6. In brief, pBAD24-mCherry-transformed *Y. pseudotuberculosis* IP2666 Δ6 was mated with *E. coli* DH5α (λ Pir) containing pSC123 in the presence of the MT607 helper strain. The mating mixtures were plated onto the LB agar containing ampicillin and kanamycin for counter selection. Single transconjugants were randomly picked into 96-well plates containing 80 μl of LB broth for culture and long-term storage. Approximately 21,000 mutant clones were collected to generate the transposon library. For the screen, 293T cells stably transfected with pNF-κB-eGFP reporter plasmid were seeded into 96-well plates and cultured overnight to reach nearly 90% confluency. The transposon mutant culture (OD_{600} 1.5–2) was added onto the cells (3 μl per well) and the infection was facilitated by centrifugation at 800*g* for 5 min. One and a half hours after infection, the culture medium was replaced with fresh DMEM containing 34 μg ml^{-1} chloramphenicol. Ten hours later, the infected cells were fixed in 4% paraformaldehyde for 30 min and then incubated in PBS (60 μl per well) for eGFP fluorescence measurement using a microplate reader. The candidate mutant strains obtained from the eGFP reporter screen were further screened using the NF-κB luciferase reporter assay.

Cell culture, luciferase reporter assay and microscopy. 293T and THP-1 cells were obtained from the American Type Culture Collection (ATCC). 293T cells were grown in DMEM supplemented with 10% (v/v) fetal bovine serum (FBS) and 2 mM L-glutamine. THP-1 cells were grown in RPMI 1640 medium containing 10% FBS and 2 mM L-glutamine. Knockout 293T cell lines were generated by using the CRISPR–Cas9 method as recently described[34] and sequences for gRNAs targeting *ALPK1* and *TIFA* were listed in Supplementary Table 2. All cells were tested for mycoplasma using the standard PCR method. Cell identity was checked frequently by their morphological features but was not authenticated by short tandem repeat (STR) profiling. For luciferase assay, the plasmids were transfected into 293T cells with the Jetprime reagents (Polyplus). Luciferase activity was determined using the dual luciferase assay kit (Promega) according to the manufacturer's instructions. For fluorescence microscopy imaging, 293T cells expressing the pNF-κB-eGFP reporter were seeded onto glass coverslips in 24-well plates and cultured for 16 h before transfection or infection (MOI 50). The cells were fixed and stained with Hoechst 33342. Fluorescence images were acquired on the Nikon A1-R or Zeiss Meta confocal microscope.

Immunoprecipitation and qRT–PCR. For immunoprecipitation, 293T cells at a confluency of 70–80% in 6-well plates were transfected with a total of 2.5 μg of indicated plasmids. Twenty-four hours after transfection, cells were treated with or without ADP-LD-Hep (100 μM) for 4 h. The cells were washed once in PBS and lysed in buffer containing 50 mM Tris-HCl (pH 7.5), 150 mM NaCl, 2 mM EDTA and 1% Triton X-100, supplemented with a protease inhibitor cocktail (Roche Molecular Biochemicals). Pre-cleared cell lysates were then subjected to anti-Flag M2 immunoprecipitation following the manufacturer's instructions. The beads were washed four times with the lysis buffer and the immunoprecipitates were subjected to standard immunoblotting analysis. qRT–PCR was performed as previously described[35]. mRNA levels of the target gene were normalized to that of *18S* rRNA. The primers used for human *IL8* are 5′-AATCTGGCAACCCTAGTCTGCTA-3′ (forward) and 5′-AAACCAAGGCACAGTGGAACA-3′ (reverse), and those for human *18S* rRNA are 5′-GACTCATTGGCCCTGTAATTGGAATGAGTC-3′ (forward) and 5′-CCAAGATCCAACTACGAGCTT-3′ (reverse), both of which are the same as previously described[5].

FACS-based genome-wide CRISPR–Cas9 screen. Human CRISPR knockout pooled gRNA plasmid library (GeCKO v2) encompassing 123,411 different gRNAs targeting 19,050 human genes was generated by the Zhang laboratory[36] and obtained from Addgene. Amplification of the library and preparation of the lentivirus were performed as recently described[34]. To perform the screen, 293T cells stably expressing Cas9 and pNF-κB-eGFP reporter were seeded in the 15-cm dish (7.5×10^6 cells per dish) and a total of 1.5×10^8 cells were infected with the gRNA lentivirus library at an MOI of 0.3. Forty-eight hours after infection, cells were re-seeded and selected with 1 μg ml^{-1} puromycin. After 6 days, 2×10^8 puromycin-resistant cells were left untreated to obtain the control sample. About 2.5×10^8 puromycin-resistant cells were treated overnight with 10 μM ADP-LD-Hep and sorted for eGFP-negative cells (~15%) on a BD Biosciences FACSAria II Flow Cytometer. After culturing for about one week, the sorted cells were re-treated with ADP-LD-Hep until the percentage of eGFP-negative cells reached more than 95%. The resulting ADP-LD-Hep unresponsive cells were cultured and treated overnight with 20 ng ml^{-1} TNF followed by FACS sorting of the eGFP-positive cells. After expansion, the TNF-responsive cells, together with the control cells, were subjected to DNA extraction[34]. Two parallel screens were performed. Amplification of the gRNA sequences was performed using a two-step PCR method similar to that described recently[34]. In the first step, twenty-four 50-μl PCR reactions (each containing 10 μg of genomic DNA) were performed with the forward primer 50bp-F (5′-CTCTTTCCCTACACGACGCTCTTCCG ATCTCTTGTGGAAAGGACGAAACA-3′) and the reverse primer 50bp-R2 (5′-GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTTCTCAAGATCTAGT TACGCC-3′); the PCR program used is 95 °C for 3 min, 18 cycles of 95 °C for 30 s, 56 °C for 20 s and 68 °C for 20 s, and a final 3-min extension at 68 °C.

Products of the first-step PCR were pooled together and used as the template for the second-step PCR. Four 50-μl PCR reactions (each containing 2.5 μl of the first-step PCR product) were performed with the forward primer Index-F (5′-AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACG-3′) and one of the reverse primers: Index-R1 (5′-CAAGCAGAAGACGGCATACGAGATCGTGATGTGACTGGAGTTC-3′) for the control sample, Index-R4 (5′-CAAGCAGAAGACGGCATACGAGATTGGTCAGTGACTGGAGTTC-3′) and Index-R8 (5′-CAAGCAGAAGACGGCATACGAGATTCAAGTGT GACTGGAGTTC-3′) for the screen samples. The PCR program used is 95 °C for 3 min, 21 cycles of 95 °C for 30 s, 56 °C for 20 s and 68 °C for 20 s, and a final 3-min extension at 68 °C. The products of the second-step PCR reactions were subjected to electrophoresis on 2% agarose gel; the 274-bp DNA bands were extracted and sequenced at the HiSeq2500 platform (Illumina). The first 19 nucleotides from each sequencing read are the gRNA sequence recovered. The frequency of each gRNA was obtained by dividing the gRNA read number by the total sample read number; the fold change was calculated by comparing the frequency of each gRNA in the screen sample with that in the control sample. The fold change ranking was obtained based on the smaller fold change value of each individual gRNA in the two parallel screens.

Recombinant protein purification. Protein expression was induced in *E. coli* BL21 (DE3) strain (wild-type or Δ*hldE*) at 20 °C for 16 h with 0.5 mM isopropyl-β-D-thiogalactopyranoside (IPTG) after OD_{600} of the bacterial culture reached 0.8. Affinity purification of MBP–ALPK1-NTD and MBP–TIFA was performed using amylose resin. His_6-tagged SUMO-GmhA, SUMO-HldE, GmhB, ALPK1-NTD, and TIFA proteins were purified using Ni-NTA agarose resin (Qiagen). The proteins were further purified by HiTrap Q HP ion-exchange chromatography and gel filtration chromatography (GE Healthcare Life Sciences). Five per cent glycerol was used throughout the purification process for TIFA protein. To obtain the ALPK1-(N+K) complex, pGEX-6p-2 vector containing human ALPK1 (1–473) and pACSUMO-ALPK1 (959–1244) were transformed into *E. coli* BL21 (DE3). Cells were grown at 37 °C in LB medium containing 100 μg ml^{-1} ampicillin and 30 μg ml^{-1} kanamycin; when OD_{600} reached 0.8, 0.5 mM IPTG was added to induce protein expression at 20 °C for 16 h. Cells were collected by centrifugation at 4,700*g*, resuspended in lysis buffer containing 20 mM Tris-HCl (pH 8.0) and 300 mM NaCl, and then lysed with an ultrasonic cell disruptor. GST–ALPK1-(N+K) complexes were purified by glutathione sepharose affinity chromatography. GST was removed by overnight digestion with the homemade HRV 3C protease at 4 °C. The ALPK1-(N+K) complex was concentrated to 0.1 mg ml^{-1} for subsequent biochemical assays. Human NMNAT1 and NMNAT3 were expressed in *E. coli* BL21 (DE3) Δ*hldE* strain, and both proteins were purified in the same way as for ALPK1-(N+K), with an additional step of Superdex G200 gel filtration chromatography (GE Healthcare Life Sciences).

To obtain proteins for structure determination, pET28-His_6–SUMO–ALPK1 (1–451 or 1–446) was transformed into *E. coli* BL21 (DE3). Proteins were expressed and purified similarly as for ALPK1-(N+K) complexes except that a Ni-Sepharose column was used for affinity purification. SUMO was removed by overnight digestion with homemade ULP1 protease at 4 °C. The untagged protein was further purified by HiTrap Q anion-exchange and Superdex G75 gel filtration chromatography. Selenomethionine-substituted (SeMet) ALPK1 (1–451) was expressed in the methionine auxotrophic *E. coli* strain B834 (DE3) and purified by the same procedure as for the native protein.

Bio-layer interferometry. The bio-layer interferometry (BLI) assay was performed on Octet RED96 System (ForteBio) at 29 °C. The running buffer was 20 mM HEPES (pH 7.5), 150 mM NaCl and 0.02% Tween-20. Ligand-free ALPK1-(N+K) complexes purified from *E. coli* BL21 (DE3) Δ*hldE* were biotinylated using EZ-Link Sulfo-NHS-LC-Biotinylation Kit (Thermo Fisher Scientific) for 30 min and then immobilized onto a Super Streptavidin (SSA) biosensor that had been equilibrated in the running buffer for 10 min. The SSA biosensor was transferred into buffer containing ADP-LD-Hep, HBP or S7P at the indicated concentration (for association) or the analyte-free buffer (for dissociation). The data were analysed and the binding constant was determined using software provided by Fortebio (Data Analysis 8.5).

Enzymatic synthesis of H1P and ADP-heptose 7-P. For enzymatic synthesis of H1P, 1 mM HBP and 10 μM GmhB protein purified from *E. coli* BL21 (DE3) Δ*hldE* were reacted in 1 ml of buffer containing 20 mM Tris-HCl (pH 8.0) and 10 mM $MgCl_2$ for 2 h at 37 °C. The reaction product H1P (anion of *m*/*z* 289) was purified by HPLC-MS. For enzymatic synthesis of ADP-heptose 7-P, GmhA and HldE proteins were purified from *E. coli* BL21 (DE3) Δ*hldE*, and the synthesis was performed with 1 mM S7P, 4 mM ATP, 10 μM GmhA, and 5 μM HldE in 20 mM HEPES (pH 7.4), 20 mM KCl and 100 mM $MgCl_2$ at 30 °C for 4 h. The reaction was stopped by incubation at 95 °C for 5 min, and then centrifuged at 13,000 rpm for 5 min to remove protein precipitates. ADP-heptose 7-P (anion of *m*/*z* 698) was purified from the reaction by HPLC-MS as described below.

HPLC-MS analysis, fractionation and LC–MS/MS. Enzymatic reactions and extracts of *E. coli*-purified His_6–ALPK1-NTD were subjected to HPLC-MS analysis or fractionation performed on Waters HPLC (Column: Atlantis T3; eluents: 0.1% NH_4HCO_3/H_2O and CH_3CN) with 2998PDA and 3100MS detectors (ESI ionization). Liquid chromatography with tandem mass spectrometry (LC–MS/MS) analysis was performed on an Agilent 1290 Infinity HPLC coupled with an Agilent 6540 quadrupole time of flight mass spectrometer. A Phenomenex Kinetex F5 column (2.1 × 100 mm, 2.6 μm) was used for separation. Mobile phases A and B were 0.1% formic acid-containing water and acetonitrile, respectively. Column temperature was set to 35 °C and the flow rate was 0.4 ml min^{-1}. The following gradient was applied: 0–4 min 0% B, 4–6.5 min from 0% to 50% B, 6.5–6.6 min from 50% to 100% B, 6.6–8 min 100% B, 8–8.2 min from 100% to 0% B, and 8.2–10 min re-equilibration at 0% B. Six microlitres of each sample was injected into the instrument and the mass spectrometry data were collected in positive and negative ionization modes for detecting ADP-Hep and ADP-heptose 7-P, respectively. A collision energy of 20 V was applied for the MS/MS acquisition of ADP-Hep.

Native mass spectrometry analysis. His_6–ALPK1-NTD protein purified from *E. coli* BL21 (DE3) was used for native-ESI analysis. Purified recombinant proteins (10 μM) were buffer-exchanged into 100 mM ammonium acetate (pH 7.5) using a centrifugal buffer exchange column (Micro Bio-Spin 6, Bio-Rad), and one aliquot was denatured by adding 0.1% formic acid (final concentration). Both the native and denatured protein samples were analysed by direct infusion. Specifically, 3 μl of each protein sample was loaded into a nano-flow capillary (borosilicate emitters, Thermo Scientific) and sprayed into a high-resolution mass spectrometer Q-Exactive HF through the nanospray FLEX ion source. The mass spectrometer settings were: 0.8–1.2 kV for capillary voltage; S-lens RF level of 100, SID (100) for complete desolvation of the native protein sample; capillary temperature at 150 °C for native protein samples or 200 °C for denatured samples; scan range 2,000–6,000 *m*/*z* for native samples and 800–6,000 for denatured samples; intact protein mode with trapping gas pressure set as 0.2. Mass spectra were analysed using the Thermo Scientific Protein Deconvolution software[37]. The parameters were specified according to the mass spectrometer settings. The minimum adjacent range of charges was 3–6 for native proteins or 7–20 for the denatured proteins, and mass tolerance was 30 p.p.m. The deconvoluted mass of the most abundant ion was selected as the mass of the target protein. The mass of the bound ligand was calculated as the difference between the native protein and the denatured protein.

Ultra-performance liquid chromatography with tandem mass spectrometry analysis of ADP-heptose 7-P generated from HBP by NMNAT1. The Dionex Ultimate 3000 UPLC system was coupled to a TSQ Quantiva Ultra triple-quadrupole mass spectrometer (Thermo Fisher), equipped with a heated electrospray ionization probe. Samples were separated using a synergi Hydro-RP column (2.0 × 100 mm, 2.5 μm, Phenomenex). A binary solvent system (mobile phase A, 10 mM tributylamine adjusted with 15 mM acetic acid in water; mobile phase B, methanol) was used. A 15-min gradient with a flow rate of 250 μl min^{-1} was applied as follows: 1–7 min at 15% B; 7–12 min, 15–98% B; 12–12.1 min, 98–15% B; 12.1–15 min, 15% B. The column chamber and sample tray were held at 45 °C and 10 °C, respectively. Data were acquired in the selected reaction monitoring mode for ADP-heptose 7-P with transitions of 698/351 in the negative ion mode. Both the precursor and fragment ions were collected at the resolution of 0.7 FWHM. The source parameters are as follows: spray voltage, 3,000 V; ion transfer tube temperature, 350 °C; vaporizer temperature, 300 °C; sheath gas flow rate, 35 Arb; auxiliary gas flow rate, 12 Arb; CID gas, 1.5 mTorr. Data analysis and quantification were performed using the software Xcalibur 3.0.63 (Thermo Fisher).

Chemical cross-linking coupled with mass spectrometry. To obtain well-behaved apo-ALPK1-(N+K) complexes for chemical cross-linking coupled with mass spectrometry analysis, GST–ALPK1-NTD was co-expressed with SUMO–ALPK1-KD in *E. coli* BL21 (DE3) Δ*hldE*. The GST–ALPK1-(N+K) complexes were immobilized on a glutathione sepharose column and washed with 20 bed volumes of 20 mM Tris-HCl (pH 8.0) containing 500 mM NaCl. GST was removed by overnight digestion with a homemade HRV 3C protease at 4 °C. The supernatant containing the released apo-ALPK1-(N+K) was passed through fresh glutathione sepharose beads, further purified by gel filtration chromatography, and concentrated to 1 mg ml^{-1}.

About 10 μg of apo-ALPK1-(N+K) complex incubated with or without 0.1 mM of the indicated sugar molecule was crosslinked with 0.5 mM disuccinimidyl suberate (DSS) at 25 °C for 40 min, and the reaction was quenched by 20 mM NH_4HCO_3. Proteins were precipitated with ice-cold acetone, resuspended in 8 M urea in 100 mM Tris-HCl (pH 8.5), digested sequentially with rLys-C (Promega) for 2 h and trypsin (in 2 M urea and 100 mM Tris-HCl, pH 8.5) for 12 h. LC–MS/MS analyses were performed on an Easy-nLC 1000 II HPLC (Thermo Fisher Scientific) coupled to a Q-Exactive HF mass spectrometer (Thermo Fisher Scientific). Peptides were loaded on a pre-column (75 μm ID, 6 cm long, packed with ODS-AQ 120 Å–10 μm beads from YMC) and further separated on an analytical column (75 μm ID, 12 cm long, packed with C18 1.8 μm 100 Å resin from Welch Materials) using a linear reverse-phase gradient from 100% buffer A (0.1% formic acid in H_2O) to 28% buffer B (0.1% formic acid in acetonitrile) in 56 min at a flow rate of 220 nl min^{-1}. The top 15 most intense precursor ions from each full scan (resolution 60,000)

were isolated for HCD MS2 (resolution 15,000; normalized collision energy 27) with a dynamic exclusion time of 45 s. Precursors with 1+, 2+, 7+ or above, or unassigned charge states were excluded. Each sample was analysed twice and the two technical repeats were combined for data analysis. pLink software[38] was used to identify cross-linked peptides with precursor and fragment ion mass accuracy at 20 p.p.m., and the results were filtered by applying a 5% FDR cutoff at the spectral level and then an *E*-value cutoff at 0.001. The inter-domain crosslinks were further filtered by requiring >3 spectra and the best *E*-value $< 1.0 \times 10^{-7}$ in at least one of the samples to eliminate false identifications[39].

In vitro ALPK1 kinase assay. Two hundred nanograms of ALPK1-(N+K) complex purified from *E. coli* BL21 (DE3) Δ*hldE* or 3×Flag-ALPK1 (full-length) immunopurified from *ALPK1*$^{-/-}$ 293T cells were mixed with 2 μg TIFA–His_6 protein (also purified from the Δ*hldE* strain) in a 50-μl reaction containing 45 mM HEPES (pH 7.4) and 4 mM $MgCl_2$. ALPK1 K/M and TIFA T9A mutant proteins were used as negative controls. ADP-LD-Hep or HBP (20 μmol) was added to test their ability to activate ALPK1. Kinase reactions were initiated by adding 100 μM ATP and allowed to proceed for 30 min or 1 h at 30 °C. To assay HBP and its activation by NMNAT-mediated modification, 5 μg of NMNAT1 or NMNAT3 protein purified from *E. coli* BL21 (DE3) Δ*hldE* were mixed with 20 μM HBP also in a 50-μl reaction in the presence of 100 μM ATP; the reaction was incubated at 30 °C for 1 h. Following addition of 200 ng of ALPK1-(N+K) and 2 μg of TIFA–His_6, the reaction was allowed to proceed for another 1 h. The reactions were stopped by adding 4×SDS loading buffer and subjected to SDS–PAGE analyses. Phosphorylated TIFA–His_6 protein was detected by immunoblotting using the anti-pT9-TIFA antibody as described above.

Protein crystallization and structure determination. Purified proteins were concentrated to 10 mg ml^{-1} for crystallization screens at 20 °C using the sitting-drop vapour diffusion method. The drop, containing 1 μl of protein solution and 1 μl of reservoir solution, was equilibrated over 100 μl reservoir solution. Initial crystallization hits of SeMet-labelled ALPK1 (1–451) and native ALPK1 (1–446) appeared from the PEG-ion Kit and the Crystal Screen Kit (Hampton Research), respectively. Qualified crystals of SeMet-labelled ALPK1 (1–451) were obtained in the reservoir solution containing 6% Tacsimate (pH 5.8) and 6.8% PEG 3550 within 1 week, and the best-diffracted crystals of native ALPK1 (1–446) were grown from the reservoir solution containing 0.1 M CH_3COONa (pH 4.0) and 1.4 M HCOONa. For data collection, the crystals were soaked in cryoprotectant solution containing the reservoir buffer supplemented with 30% ethylene glycol for SeMet-labelled ALPK1 (1–451) or 15% glycerol for native ALPK1 (1–446) followed by flash-freezing with liquid nitrogen. Diffraction data were collected at the Shanghai Synchrotron Radiation Facility (Shanghai) beamline BL18U1 for SeMet-labelled ALPK1 (1–451) and BL19U1 for native ALPK1 (1–446) under the wavelengths of 0.97776 Å and 0.97853Å, respectively. Data were processed in X-ray Detector Software. The phase was determined by the single wavelength anomalous dispersion method and automatic model building was performed in PHENIX. The rest of the model was manually built with Coot. The structure of ALPK1-NTD (residues 1–446) was refined in PHENIX, and manual modelling was performed between refinement cycles. The statistics of data collection and refinement are summarized in Extended Data Table 1a. Ramachandran statistics indicated that all the residues are in the allowed region, and 98.2% fall into the favoured region. The quality of the final model was validated by MolProbity.

Mouse experiments and measurements of cytokines. Wild-type C57BL/6 mice were purchased from Vital River Laboratory Animal Technology (Beijing). To generate *Alpk1*$^{-/-}$ mice, a gRNA (GGCCCTTCGTGCCTGAAAAG) targeting exon 3 of *Alpk1* was designed with an online gRNA designing tool (http://crispr.mit.edu/). In vitro transcribed guide RNA and Cas9 mRNA were co-microinjected into C57BL/6 mice-derived zygotes. The tail-end genomic DNA of each offspring was amplified with the forward primer 5′-CCTGTAGGGCAGAGTAGGCT-3′ and the reverse primer 5′-TTCAAGGTGACAGGTTTCGT-3′. Sanger sequencing was performed to analyse the PCR products and identify the founders with out-of-frame indels. Founders with the same out-of-frame indels were intercrossed to obtain homozygous *Alpk1*$^{-/-}$ mice (Supplementary Table 2). All mice were maintained in the specific pathogen-free facility at National Institute of Biological Sciences, Beijing. All mouse experiments were carried out in accordance with the national guidelines for housing and care of laboratory animals (Ministry of Health, China) and the protocol is in accordance with institutional regulations after review and approval by the Institutional Animal Care and Use Committee at National Institute of Biological Sciences, Beijing.

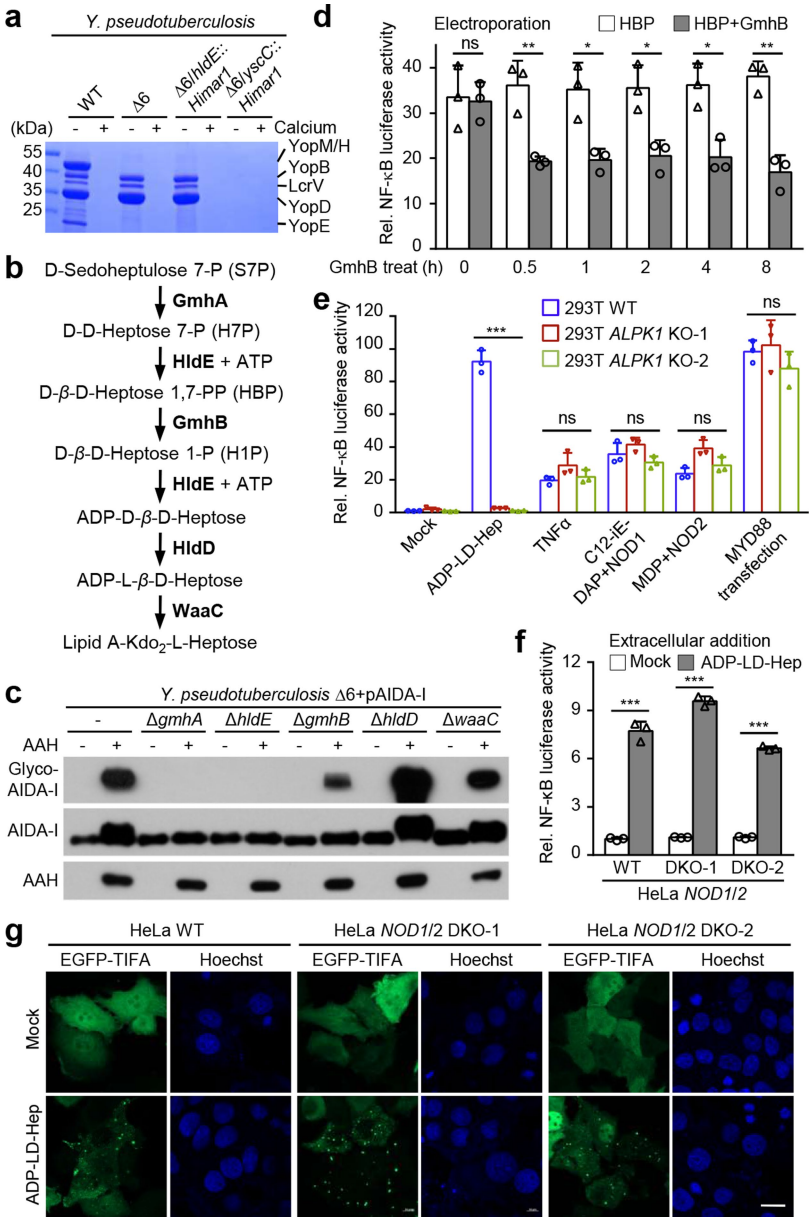
Eight-week-old female wild-type or *Alpk1*$^{-/-}$ C57BL/6 mice were used for air pouch model construction, followed by HBP or ADP-LD-Hep administration. Five mice were assayed for each group. The air pouch was constructed as previously described[40], with slight modifications. In brief, 3 ml of sterile air was injected into the subcutaneous tissue of the back of the mice. Three days later, another 2.5 ml sterile air was injected into the same pouch to maintain patency. Six days later, synthetic HBP or ADP-LD-Hep (2 mg kg^{-1} body weight) dissolved in 300 μl of saline was injected into the well-developed pouch (saline alone was injected into the control mice). Another 3 h later, blood samples from mouse tail end were collected for serum isolation, the mice were then killed, and the pouch was opened with a small hole. Immediately afterwards, 600 μl of PBS was injected into the pouch, and the fluids were gently sucked into and out of the bulb to mix the contents. All fluids within the pouch were collected and centrifuged; the supernatants and the cell pellets were used for ELISA analysis and counting the total leukocyte number, respectively. Cytokine levels in the sera and air pouch fluid supernatants were determined using ELISA kits (IL-6, IP-10, GRO-α and MCP1). ProcartaPlex multiplex immunoassay (eBioscience) was also performed to measure up to 36 cytokines in both the air pouch washes and the sera.

For infection, single clones of *B. cenocepacia* J2315 strain[30] were cultured at 37 °C in LB broth with shaking for 18 h. The bacterial cultures were diluted by 1:50 in fresh LB broth, and grown until OD_{600} reached 0.8. The bacteria were washed with PBS and then diluted in 1% gelatin-containing PBS. Indicated 6-week-old male mice (C57BL/6 background) were anaesthetized by intraperitoneal injection of 0.8% pentobarbital sodium and infected intratracheally with 30 μl of bacteria suspension (5×10^7 c.f.u. for each mouse) or 1% gelatin-containing PBS. The mice were killed 24 h post-infection. The lungs were removed and homogenized to measure bacterial burden or cytokine concentration by ProcartaPlex multiplex immunoassay (eBioscience Inc.).

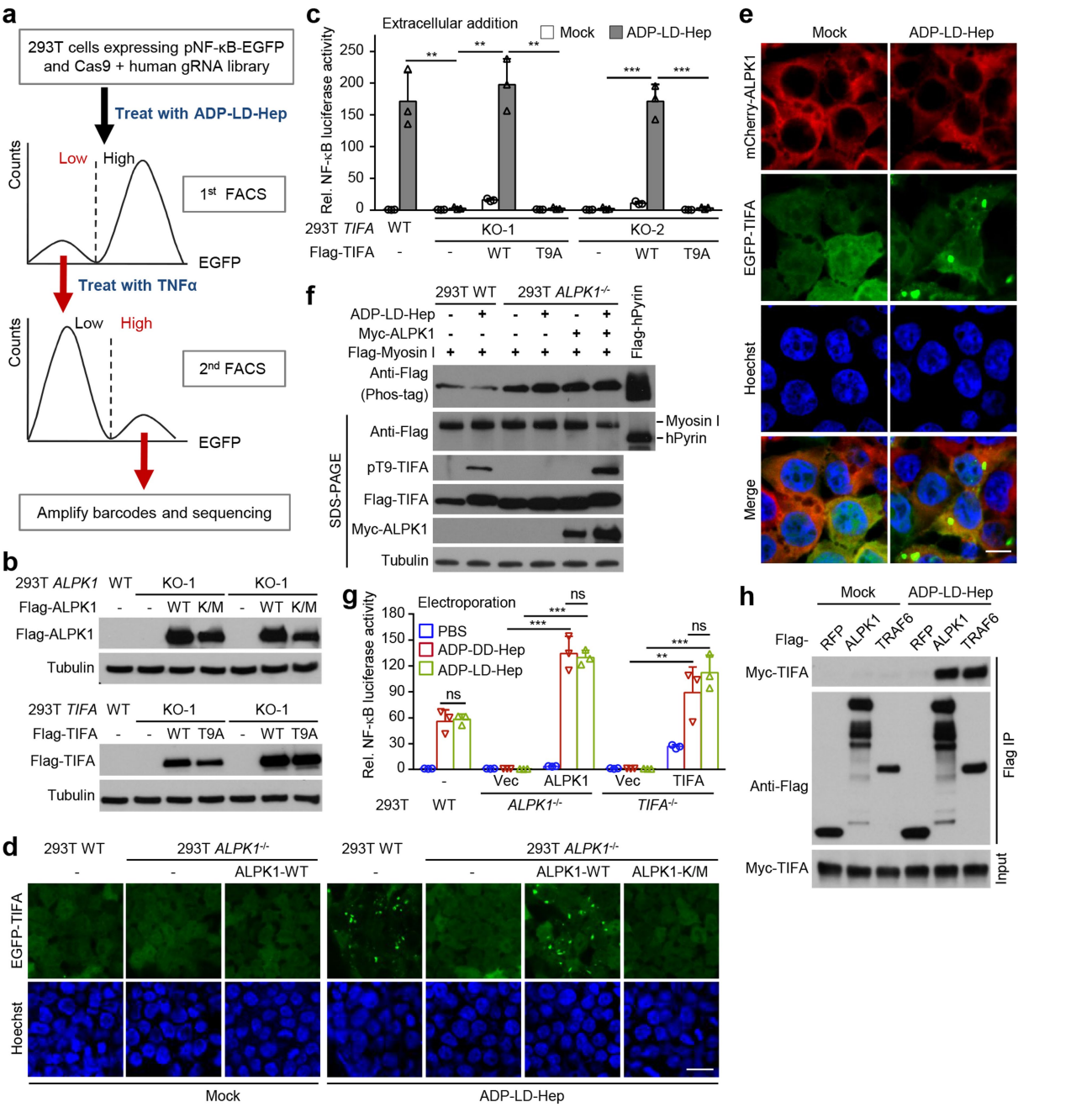
Reporting summary. Further information on experimental design is available in the Nature Research Reporting Summary linked to this paper.

Data availability. All data supporting the findings of this study are included in this manuscript and its Supplementary Information files. The atomic coordinates and structure factors of the ALPK1-NTD–ADP-heptose complex have been deposited in the Protein Data Bank under the accession code 5Z2C.

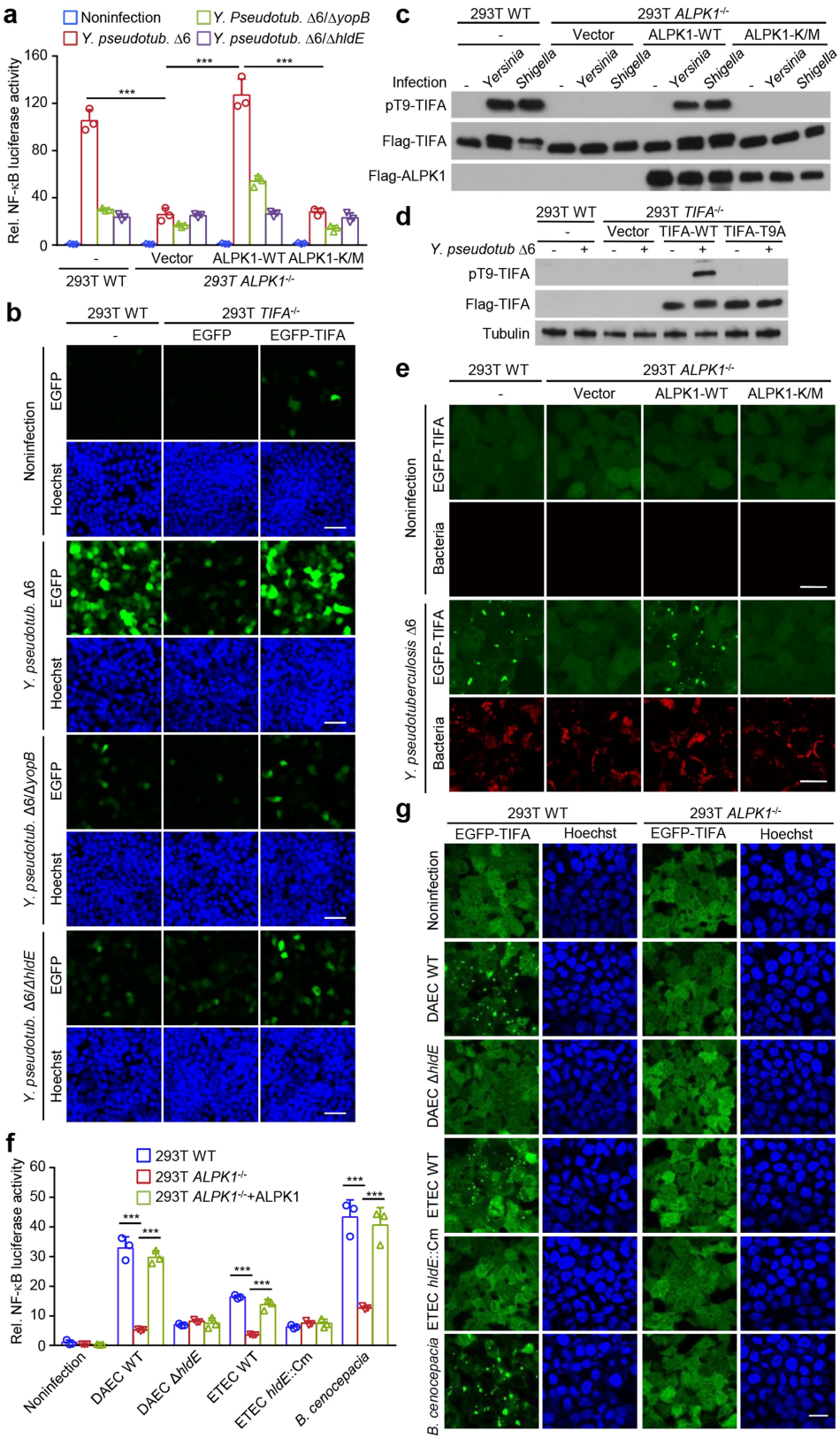
23. Ge, J. et al. A Legionella type IV effector activates the NF-κB pathway by phosphorylating the IκB family of inhibitors. *Proc. Natl Acad. Sci. USA* **106**, 13725–13730 (2009).
24. Li, H. et al. The phosphothreonine lyase activity of a bacterial type III effector family. *Science* **315**, 1000–1003 (2007).
25. Zhang, L. et al. Cysteine methylation disrupts ubiquitin-chain sensing in NF-κB activation. *Nature* **481**, 204–208 (2011).
26. Li, S. et al. Pathogen blocks host death receptor signalling by arginine GlcNAcylation of death domains. *Nature* **501**, 242–246 (2013).
27. Zamyatina, A. et al. Efficient chemical synthesis of both anomers of ADP L-glycero- and D-glycero-D-manno-heptopyranose. *Carbohydr. Res.* **338**, 2571–2589 (2003).
28. Borio, A., Hofinger, A., Kosma, P. & Zamyatina, A. Chemical synthesis of the innate immune modulator-bacterial D-glycero-β-D-manno-heptose-1,7-bisphosphate (HBP). *Tetrahedr. Lett.* **58**, 2826–2829 (2017).
29. Dong, N. et al. Structurally distinct bacterial TBC-like GAPs link Arf GTPase to Rab1 inactivation to counteract host defenses. *Cell* **150**, 1029–1041 (2012).
30. Xu, H. et al. Innate immune sensing of bacterial modifications of Rho GTPases by the Pyrin inflammasome. *Nature* **513**, 237–241 (2014).
31. Murphy, K. C. & Campellone, K. G. Lambda Red-mediated recombinogenic engineering of enterohemorrhagic and enteropathogenic *E. coli*. *BMC Mol. Biol.* **4**, 11 (2003).
32. Sorg, J. A., Miller, N. C., Marketon, M. M. & Schneewind, O. Rejection of impassable substrates by *Yersinia* type III secretion machines. *J. Bacteriol.* **187**, 7090–7102 (2005).
33. Rietsch, A., Wolfgang, M. C. & Mekalanos, J. J. Effect of metabolic imbalance on expression of type III secretion genes in *Pseudomonas aeruginosa*. *Infect. Immun.* **72**, 1383–1390 (2004).
34. Shi, J. et al. Cleavage of GSDMD by inflammatory caspases determines pyroptotic cell death. *Nature* **526**, 660–665 (2015).
35. Zhao, Y. et al. The NLRC4 inflammasome receptors for bacterial flagellin and type III secretion apparatus. *Nature* **477**, 596–600 (2011).
36. Sanjana, N. E., Shalem, O. & Zhang, F. Improved vectors and genome-wide libraries for CRISPR screening. *Nat. Methods* **11**, 783–784 (2014).
37. Hao, Z. et al. In *State-of-the-Art and Emerging Technologies for Therapeutic Monoclonal Antibody Characterization Volume 3. Defining the Next Generation of Analytical and Biophysical Techniques* (eds. Schiel, J. E. et al.) Ch. 10, 289–315 (American Chemical Society, 2015).
38. Yang, B. et al. Identification of cross-linked peptides from complex samples. *Nat. Methods* **9**, 904–906 (2012).
39. Ding, Y. H. et al. Modeling protein excited-state structures from "over-length" chemical cross-links. *J. Biol. Chem.* **292**, 1187–1196 (2017).
40. Sin, Y. M., Sedgwick, A. D., Chea, E. P. & Willoughby, D. A. Mast cells in newly formed lining tissue during acute inflammation: a six day air pouch model in the mouse. *Ann. Rheum. Dis.* **45**, 873–877 (1986).



Extended Data Fig. 1 | Analyses of the ADP-Hep biosynthesis pathway and ADP-Hep-induced NF-κB activation. a, Low calcium-induced secretion of the T3SS substrates in the *hldE* transposon mutant. *yscC* is required for T3SS assembly. **b**, Schematic of the classical ADP-Hep biosynthesis pathway in Gram-negative bacteria. **c**, ADP-Hep-dependent autotransporter heptosylation in Δ*gmhB* mutant. A fragment of the AIDA-I autotransporter (GST–AIDA-$I_{50–600}$–Flag) and its heptosyltransferase (AAH) were expressed in *Y. pseudotuberculosis* Δ6 deleted of a gene in the ADP-Hep biosynthesis pathway. AIDA-I heptosylation was assessed using the ECL glycoprotein detection kit. **d**, NF-κB activation in 293T cells electroporated with HBP or H1P (obtained by treating HBP with recombinant GmhB for the indicated times). **e**, Effects of *ALPK1* knockout on other known NF-κB pathways. NOD1, NOD2 and MYD88 were transfected into the cells. Wild-type cells and two *ALPK1*$^{-/-}$ 293T clones (KO-1/2) were treated with ADP-LD-Hep (100 μM), TNF (20 ng ml^{-1}), C12-iE-DAP (10 ng ml^{-1}) or MDP (10 ng ml^{-1}). **f**, **g**, Effects of NOD1/2 deficiency on ADP-Hep-induced NF-κB activation and TIFA foci formation. Wild-type HeLa cells and two NOD1/2 double-knockout clones (DKO-1/2) were assayed. eGFP–TIFA was transfected into the cells (**g**). Scale bar, 20 μm. **d**–**f**, NF-κB activation was assessed by luciferase reporter assay (mean ± s.d. from three technical replicates); two-tailed unpaired Student's *t*-test was performed (*$P < 0.05$, **$P < 0.01$, *** $P < 0.001$, NS, not significant). All data are representative of three independent experiments.

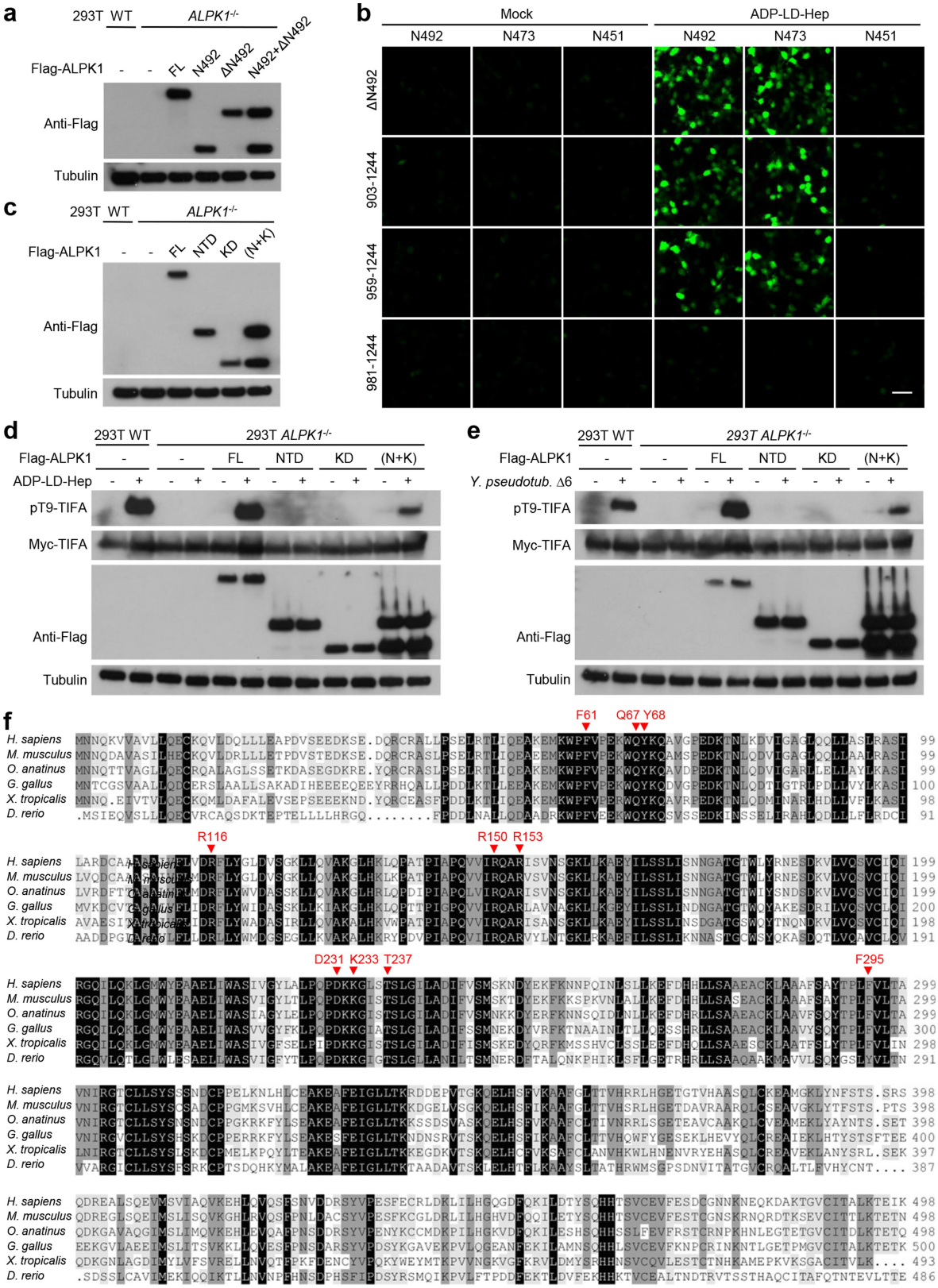


Extended Data Fig. 2 | FACS-based genome-wide CRISPR–Cas9 screen identifies the ALPK1–TIFA–TRAF6 axis that mediates ADP-Hep-induced NF-κB activation. a, Flow chart of the CRISPR–Cas9 screen for genes required for activation of NF-κB induced by extracellular ADP-LD-Hep but not TNF in 293T cells. **b**, Immunoblots of Flag–ALPK1 (wild type or its K/M mutant) and Flag–TIFA (wild type or its T9A mutant) expressed in knockout 293T cells. **c**, Requirement of T9 of TIFA for ADP-Hep-induced NF-κB activation. Flag-TIFA (wild type or T9A) was transfected into two *TIFA*$^{-/-}$ 293T clones (KO-1/2). **d**, Requirement of ALPK1 kinase activity for ADP-LD-Hep-induced formation of TIFA foci. eGFP–TIFA was stably expressed in wild-type or *ALPK1*$^{-/-}$ 293T cells expressing Flag–ALPK1 (wild type or the kinase-dead K/M mutant). Scale bar, 20 μm. **e**, Fluorescence imaging of mCherry–ALPK1 and eGFP–TIFA in 293T cells treated with or without ADP-LD-Hep. Scale bar, 10 μm. **f**, Phos-tag gel assay of myosin I phosphorylation in ADP-LD-Hep-treated cells. Flag–myosin I was transfected into indicated 293T cells. Immunoblots of total cell lysates separated on phos-tag gel or regular SDS gels are shown. **g**, Effects of *ALPK1* or *TIFA* knockout on activation of NF-κB induced by ADP-LD or DD-Hep electroporation. *ALPK1*$^{-/-}$ and *TIFA*$^{-/-}$ 293T cells were complemented as indicated. **h**, ADP-LD-Hep-induced co-immunoprecipitation of TIFA with ALPK1 and TRAF6 in transfected 293T cells. **c**, **g**, NF-κB activation was assessed by luciferase reporter assay (mean ± s.d. from three technical replicates); two-tailed unpaired Student's *t*-test was performed (*$P < 0.05$, **$P < 0.01$, ***$P < 0.001$, NS, not significant). **d**, **e**, Confocal images with Hoechst-stained nuclei. All data are representative of three independent experiments.

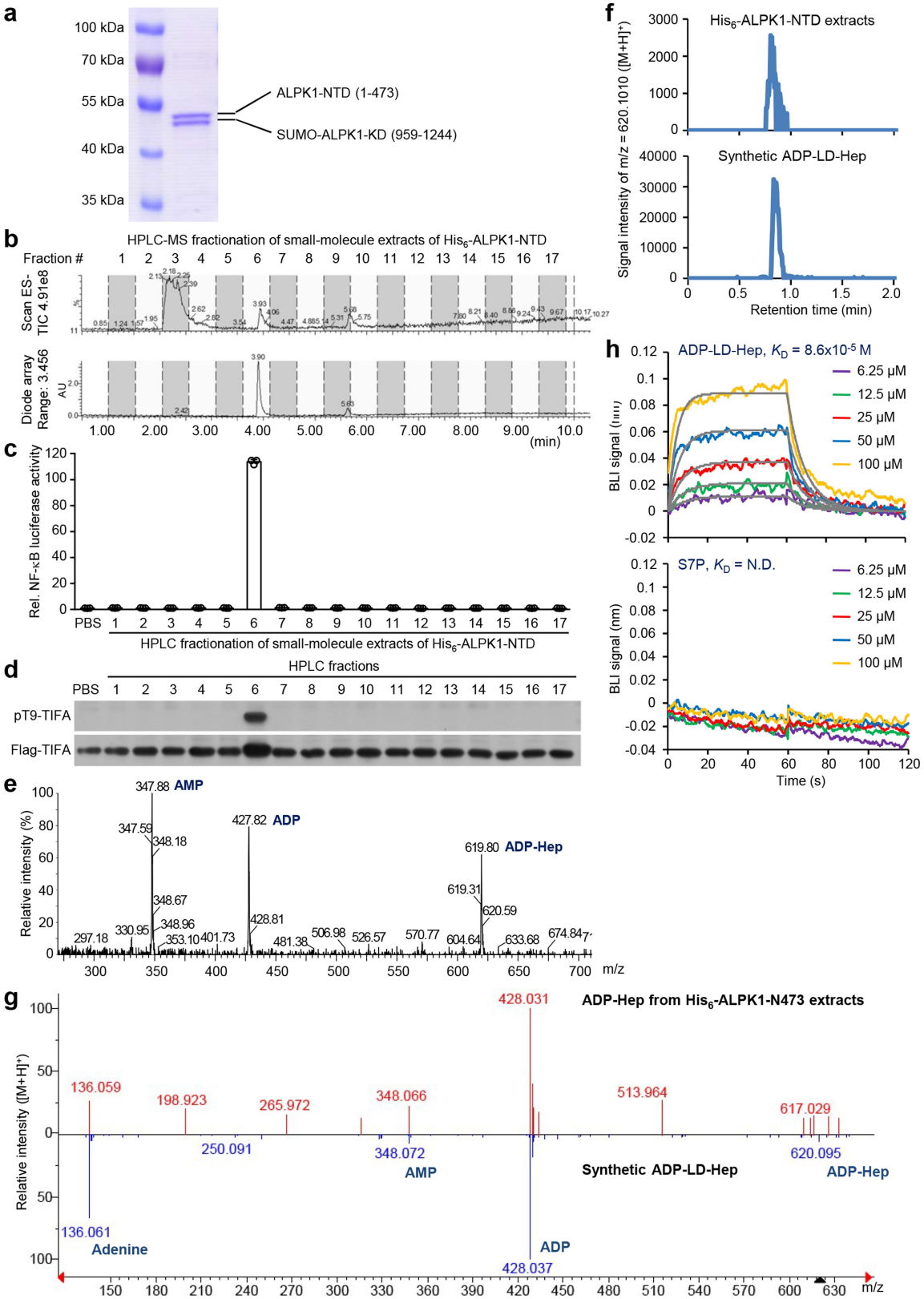


Extended Data Fig. 3 | See next page for caption

Extended Data Fig. 3 | The ALPK1–TIFA pathway mediates T3SS-dependent and -independent activation of NF-κB by various bacterial pathogens. a, **b**, Requirement of ALPK1 kinase activity and TIFA for *Y. pseudotuberculosis*-induced activation of NF-κB. Wild-type or *ALPK1*$^{-/-}$ 293T cells expressing Flag–ALPK1 (wild type or K/M) or *TIFA*$^{-/-}$ 293T cells expressing Flag–TIFA were infected with indicated *Y. pseudotuberculosis* strains. NF-kB activation was assessed by luciferase reporter assay (**a**) or eGFP reporter expression (**b**). Scale bar, 50 μm (**b**). **c**–**e**, Requirement of ALPK1 kinase activity for *Y. pseudotuberculosis* (or *S. flexneri* 2457T)-induced phosphorylation of TIFA (**c**, **d**) or formation of eGFP–TIFA foci (**e**). Flag– or eGFP–TIFA was expressed in wild-type or *ALPK1*$^{-/-}$ 293T cells expressing Flag–ALPK1 (wild type or K/M). TIFA phosphorylation was assessed by anti-pT9-TIFA immunoblotting (**c**, **d**). *Y. pseudotuberculosis* Δ6 was labelled with mCherry (scale bar, 20 μm) (**e**). **f**, **g**, NF-κB luciferase reporter activation and formation of eGFP–TIFA foci induced by T3SS-negative bacteria. Wild-type, *ALPK1*$^{-/-}$ or Flag–ALPK1-complemented *ALPK1*$^{-/-}$ 293T cells were infected with DAEC 2787, ETEC H10407, *B. cenocepacia* J2315 or their Δ*hldE* mutants. Scale bar, 20 μm. **a**, **f**, Luciferase data are shown as mean ± s.d. from three technical replicates; two-tailed unpaired Student's *t*-test was performed (*$P < 0.05$, **$P < 0.01$, *** $P < 0.001$). **b**, **e**, **g**, Confocal images with Hoechst-stained nuclei. All data shown are representative of three independent experiments.

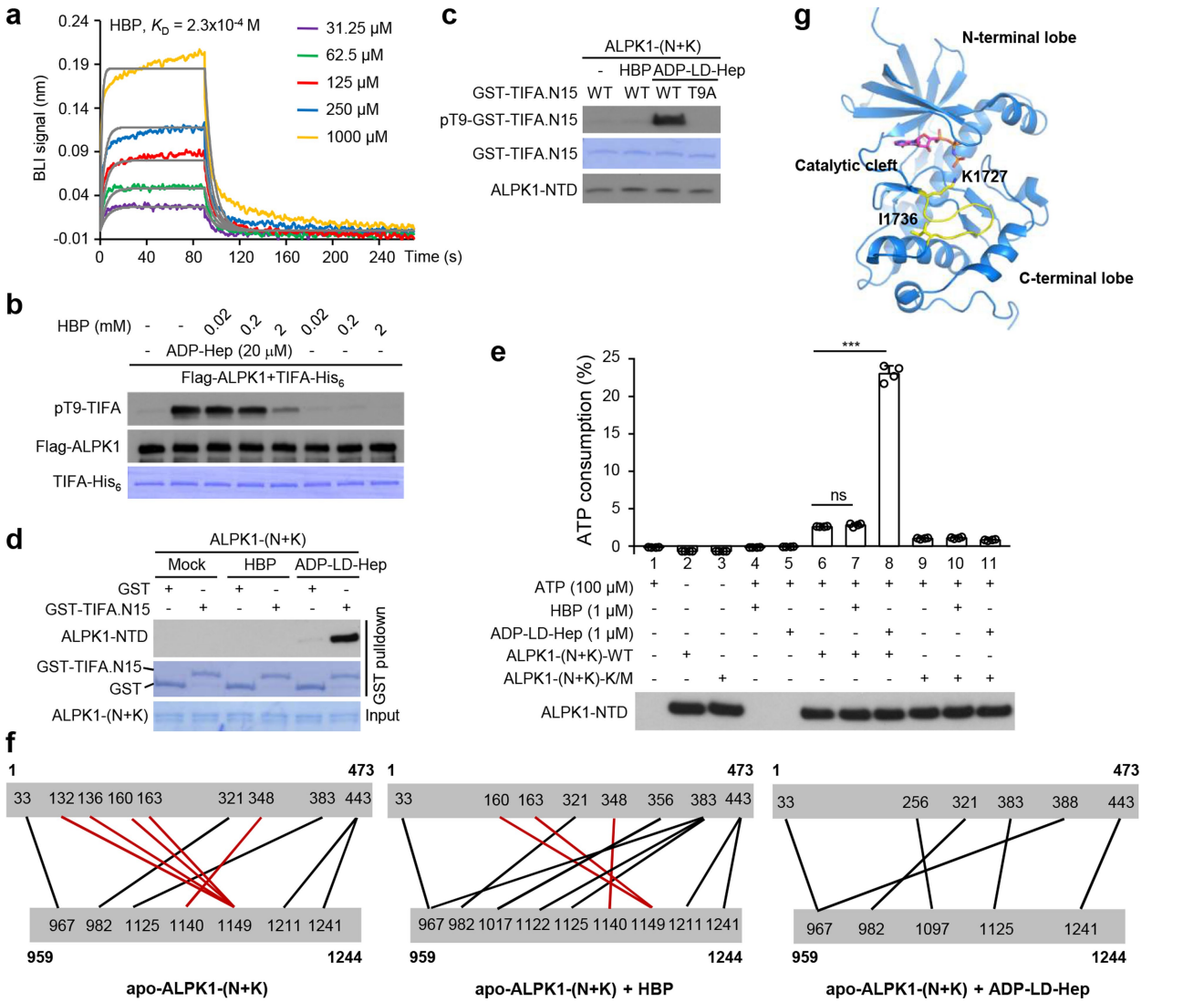


Extended Data Fig. 4 | Co-expression of ALPK1-NTD and ALPK1-KD can support ADP-Hep and *Y. pseudotuberculosis*-induced activation of NF-κB and conservation of ADP-Hep-binding residues in ALPK1-NTD. a, **c**, Immunoblots of ALPK1 N- and C-terminal truncation mutants expression in indicated 293T cells. NF-κB luciferase reporter activation induced by ADP-LD-Hep and *Y. pseudotuberculosis* Δ6 are in Fig. 2h. **b**, Mapping the minimal N- and C-terminal regions of ALPK1 sufficient for ADP-LD-Hep-induced activation of NF-κB. Fluorescence images of NF-κB–eGFP expression are shown. Scale bar, 50 μm. **d**, **e**, Assay of ALPK1-NTD and ALPK1-KD co-expression in mediating ADP-LD-Hep (**d**) or *Y. pseudotuberculosis* (**e**)-induced TIFA phosphorylation. Immunoblots of total cell lysates using indicated antibodies are shown. **f**, Multiple sequence alignment of the NTDs of ALPK1 from indicated organisms. Red residues are those involved in ADP-Hep binding, revealed by the crystal structure of human ALPK1-NTD–ADP-LD-Hep complex (Fig. 3g). Data shown are representative of three (**a**–**c**) or two (**d**, **e**) independent experiments.

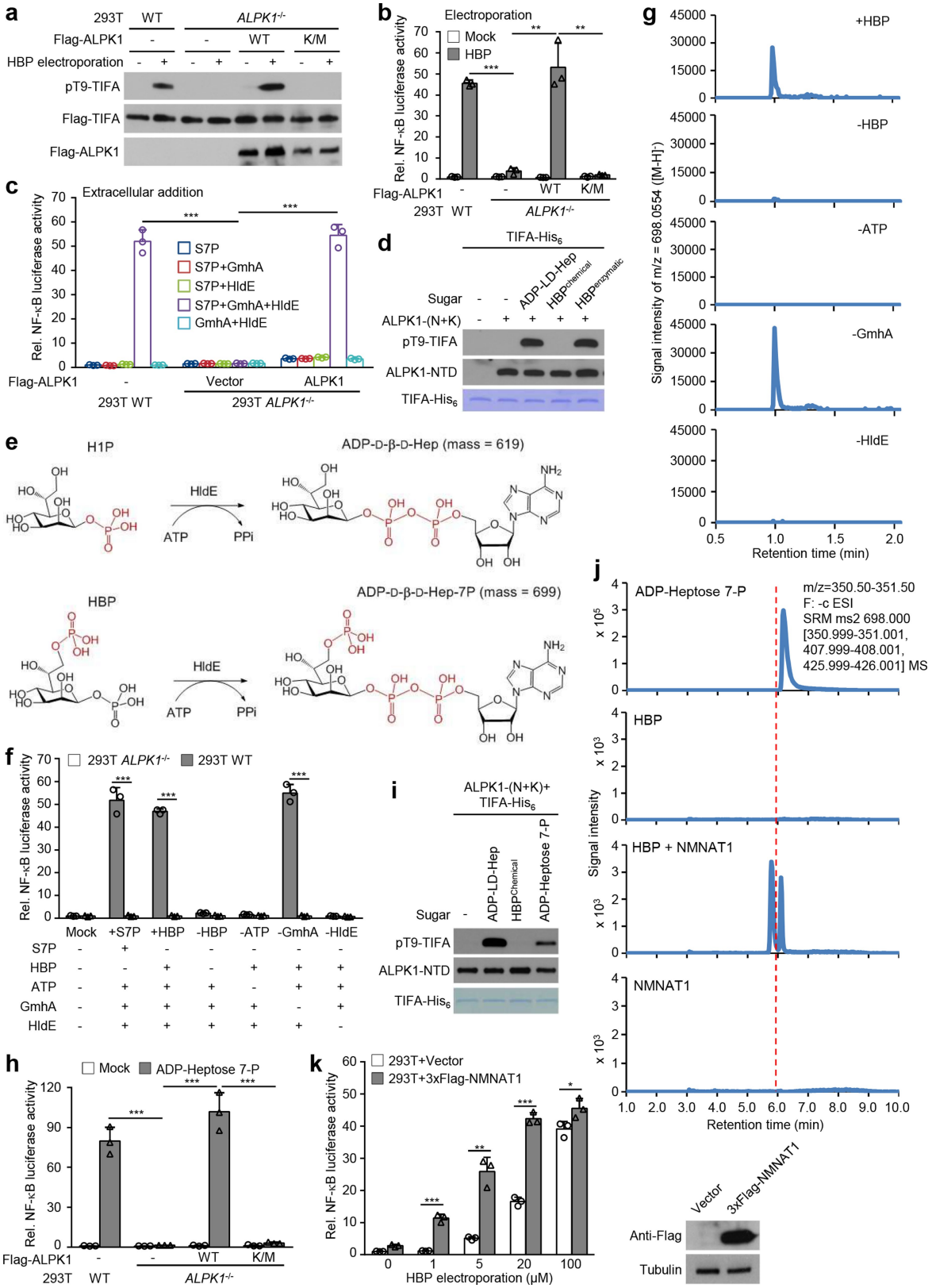


Extended Data Fig. 5 | See next page for caption

Extended Data Fig. 5 | Direct binding of ADP-Hep to ALPK1-NTD. **a**, *E. coli*-purified ALPK1-(N+K) complex on a Coomassie blue-stained SDS–PAGE gel. **b**–**e**, HPLC–MS fractionation, NF-κB-inducing activity and mass spectrometry of small-molecule extracts from His_6–ALPK1-NTD purified from wild-type *E. coli* BL21 (DE3). The small-molecule extracts, obtained by protein denaturing and precipitation (95 °C for 5 min), were analysed by HPLC–MS with 17 fractions obtained (**b**). The 17 fractions were used to treat 293T cells (**c**) or cells expressing Flag–TIFA (**d**); NF-κB luciferase activity (mean ± s.d. from three technical replicates) and anti-pT9-TIFA immunoblotting are in **c** and **d**, respectively. **e**, Mass spectrometry of fraction 6 identified three major ions corresponding to AMP, ADP and ADP-Hep. **f**, LC–MS/MS of ADP-Hep in fraction 6 (**b**, **e**) or synthetic ADP-LD-Hep standard. **g**, MS/MS spectra of the $[M+H]^+$ product ions of ADP-Hep in fraction 6 in comparison with those of synthetic ADP-LD-Hep. The heptose of ADP-Hep was not shown owing to neutral loss. **h**, BLI assay of ADP-LD-Hep or S7P binding to ALPK1-(N+K). ALPK1-(N+K) complexes purified from *E. coli* BL21 (DE3) Δ*hldE* were biotinylated in vitro. Sensorgrams of the binding to ALPK1-(N+K) by different concentrations of the indicated sugar (colour lines) are shown. Grey lines are from model fits. Data shown are representative of three (**a**, **h**) or two (**b**–**g**) independent experiments.

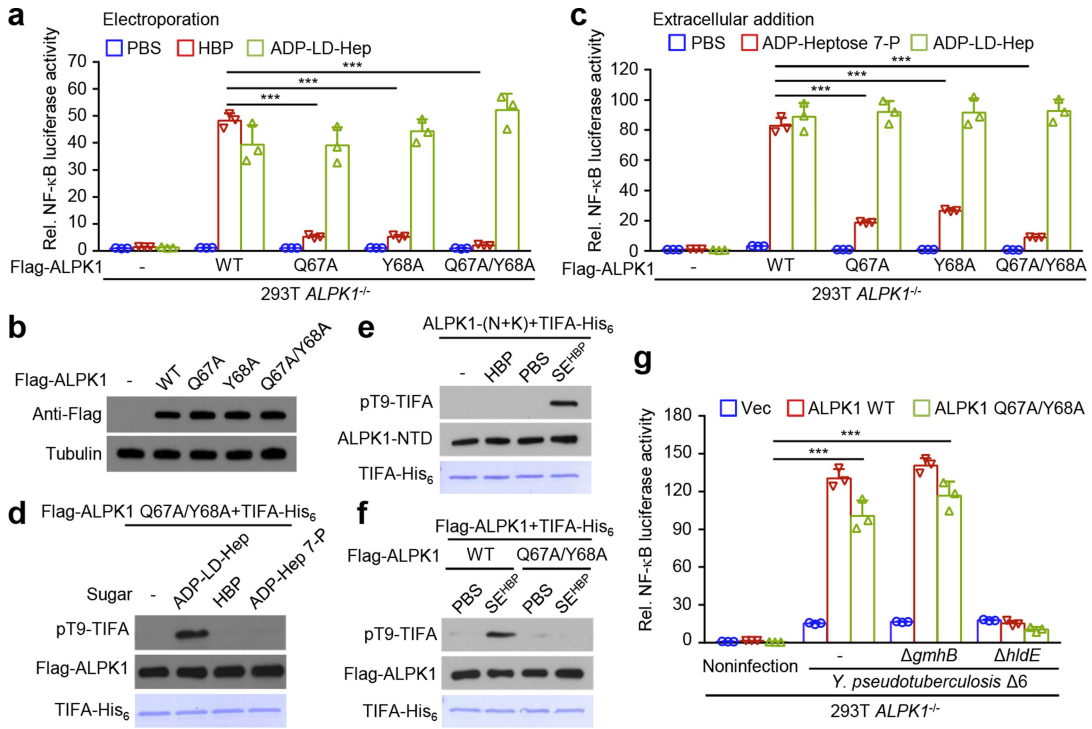


Extended Data Fig. 6 | HBP binding is insufficient to render ALPK1-KD competent for substrate recognition and phosphate transfer. a, BLI assay of HBP binding to in vitro-biotinylated apo-ALPK1-(N+K). Sensorgrams of the binding in different concentrations of HBP (colour lines) are shown. Grey lines are from model fits. **b**, Excess HBP competitively inhibits activation of ALPK1 by ADP-Hep. Flag–ALPK1 purified from 293T cells was incubated with purified TIFA–His$_6$ and ADP-Hep in the presence of a titrating amount of HBP. **c**, **d**, Effects of HBP binding on ALPK1 phosphorylation and recognition of a 15-residue peptide substrate derived from TIFA. The N-terminal 15-residue sequences of TIFA were fused to GST (GST–TIFA.N15). **d**, GST-pulldown assay of ALPK1-(N+K) binding to GST–TIFA.N15. **e**, Effect of HBP binding on ATP hydrolysis activity of ALPK1. Apo-ALPK1-(N+K) (wild type or K/M) was incubated with HBP or ADP-LD-Hep, and further reacted with ATP. Percentages of ATP consumption at indicated reaction conditions are shown (mean ± s.d. from three technical replicates). Two-tailed unpaired Student's *t*-test was performed (*$P < 0.05$; **$P < 0.01$; ***$P < 0.001$). **f**, Graphical representation of DSS-crosslinked residues between ALPK1-NTD and ALPK1-KD identified by chemical cross-linking coupled with mass spectrometry. Apo-ALPK1-(N+K) was left untreated, or incubated with HBP or ADP-LD-Hep. Crosslinking connections are depicted by straight lines, and the corresponding raw mass spectrometry data are in Supplementary Table 3. **g**, Crystal structure of TRPM7 kinase domain in complex with AMPPNP (PDB code, 1IA9). TRPM7 is shown in cartoons and AMPPNP is in sticks. The loop containing K1727 and I1736 (shown in sticks, corresponding to K1140 and K1149 in human ALPK1, respectively) is in yellow. Phosphorylation of TIFA and GST-TIFA.N15 was assayed by anti-pT9-TIFA immunoblotting (**b**, **c**). Apo-ALPK1-(N+K), TIFA–His$_6$, and GST–TIFA.N15 were purified from *E. coli* BL21 (DE3) Δ*hldE* (**a**–**f**). Data shown are representative of three (**a**–**e**) and two (**f**) independent experiments.

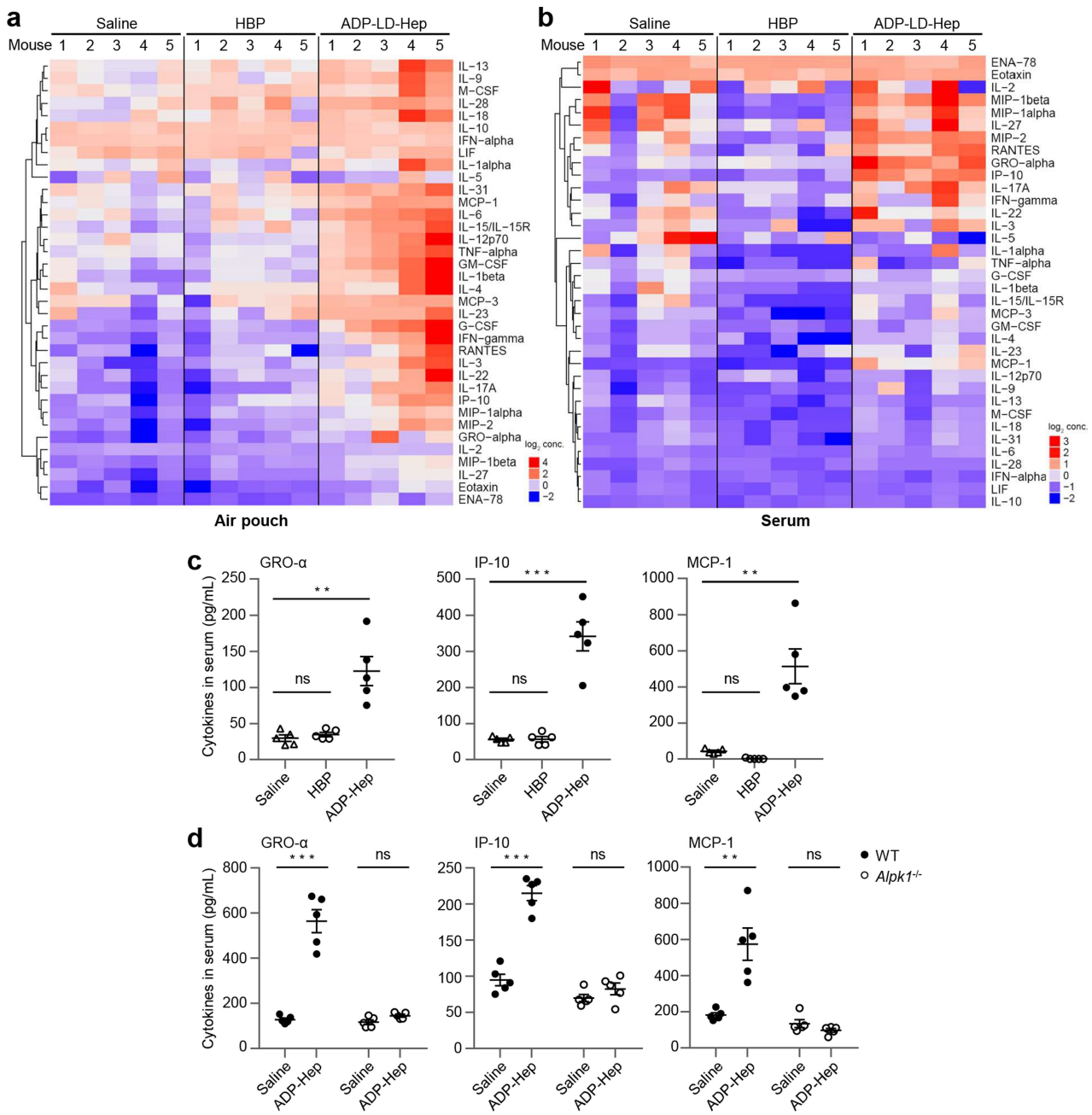


Extended Data Fig. 7 | See next page for caption

Extended Data Fig. 7 | HBP can be transformed into ALPK1 activation-competent ADP-heptose 7-P by bacterial or host adenylyltransferase. **a**, **b**, Requirement of ALPK1 kinase activity for cytosolic HBP-induced phosphorylation of TIFA (**a**) and activation of NF-κB (**b**). HBP was electroporated into wild-type or *ALPK1*$^{-/-}$ 293T cells expressing Flag–ALPK1 (wild-type or K/M). **c**, **d**, Enzymatically synthesized HBP could induce ALPK1-mediated NF-κB activation (**c**) and TIFA phosphorylation in vitro (**d**). S7P was reacted with recombinant GmhA or HldE or both. Following protein denaturing and precipitation, reaction supernatants were added to wild-type or *ALPK1*$^{-/-}$ 293T cells containing an empty vector or Flag–ALPK1 (**c**). Enzymatically synthesized HBP product (HBP^{enzymatic}), synthetic ADP-LD-Hep or HBP was incubated with ALPK1-(N+K) in the presence of TIFA–His_6 (**d**). **e**, Schematic of HldE adenylyltransferase synthesis of ADP-Hep and ADP-heptose 7-P from H1P and HBP, respectively. **f**, **g**, NF-κB activation by (**f**) and LC–MS of (**g**) enzymatically synthesized HBP product. Indicated reaction products were added to wild-type or *ALPK1*$^{-/-}$ 293T cells (**f**) or analysed by LC–MS (**g**). **h**, **i**, NF-κB (**h**) and in vitro ALPK1 activation (**i**) by ADP-heptose 7-P. HPLC-purified ADP-heptose 7-P was added to wild-type or *ALPK1*$^{-/-}$ 293T cells expressing Flag–ALPK1 (wild-type or K/M). **i**, Equal amounts of the indicated sugars were incubated with ALPK1-(N+K) in the presence of TIFA–His_6. **j**, Ultra-performance liquid chromatography with tandem mass spectrometry of the reaction product of HBP and NMNAT1. Purified ADP-heptose 7-P, synthesized by reacting HBP with HldE, was used as the standard. **k**, Effect of NMNAT1 overexpression on activation of NF-κB by electroporation of HBP into 293T cells. The immunoblots show NMNAT1 expression. Apo-ALPK1-(N+K) was purified from *E. coli* BL21 (DE3) Δ*hldE* and phosphorylation of TIFA was assessed by anti-pT9-TIFA immunoblotting (**d**, **i**). **b**, **c**, **f**, **h**, **k**, NF-κB activation was assessed by luciferase reporter assay (mean ± s.d. from three technical replicates); two-tailed unpaired Student's *t*-test was performed (*$P < 0.05$; **$P < 0.01$; ***$P < 0.001$). Data shown are representative of three (**a**–**d**, **h**, **i**, **k**) or two (**f**, **g**, **j**) independent experiments.



Extended Data Fig. 8 | ALPK1 Q67A/Y68A mutants that can discriminate cytosolic HBP from ADP-Hep resist activation by ADP-heptose 7-P. a–**c**, Effects of ALPK1 Q67A, Y68A or Q67A/Y68A double mutations on activation of NF-κB by HBP, ADP-Hep and ADP-heptose 7-P. *ALPK1*$^{-/-}$ 293T cells expressing the indicated Flag–ALPK1 mutants were electroporated with excess HBP or ADP-LD-Hep (**a**), or treated with excess ADP-heptose 7-P or ADP-LD-Hep (**c**). Anti-Flag immunoblots (**b**) show expression of the ALPK1 mutants. **d**, In vitro TIFA phosphorylation assay of ALPK1 Q67A/Y68A activation by HBP, ADP-heptose 7-P, or ADP-LD-Hep. Flag–ALPK1 Q67A/Y68A was purified from 293T cells. **e**, **f**, In vitro TIFA phosphorylation assay of ALPK1 Q67A/Y68A activation by small-molecule extracts from HBP-electroporated cells (SE^{HBP}). 293T cells electroporated with HBP were used to prepare the small-molecule extracts, and synthetic HBP was included as the control. Apo-ALPK1-(N+K) (**e**) and Flag–ALPK1 (**f**) were purified from *E. coli* BL21 (DE3) Δ*hldE* and 293T cells, respectively. **g**, Effects of ALPK1 Q67A/Y68A double mutations on *Y. pseudotuberculosis*-induced activation of NF-κB. *ALPK1*$^{-/-}$ 293T cells expressing Flag–ALPK1 (wild-type or the Q67A/Q68A mutant) were left uninfected or infected with indicated *Y. pseudotuberculosis* strains. **a**, **c**, **g**, NF-κB activation was assessed by luciferase reporter assay (mean ± s.d. from three technical replicates); two-tailed unpaired Student's *t*-test was performed (*$P < 0.05$; **$P < 0.01$; ***$P < 0.001$). All data are representative of at least two independent experiments.



Extended Data Fig. 9 | ADP-Hep but not HBP induces *Alpk1*-dependent cytokine expression in mice. HBP or ADP-LD-Hep (2 mg kg^{-1}) or the saline control were injected into the dorsal air pouches of wild-type (**a**–**d**) or *Alpk1*$^{-/-}$ (**d**) mice (C57BL/6). *n* (biologically independent animals) = 5 for each group of treatment. Cytokine profiling was determined by the multiplex immunoassay (**a**–**c**) or ELISA (**d**). **a**, **b**, Heat maps of indicated cytokine concentrations in the air pouch (**a**) and the serum (**b**) of injected mice. **c**, **d**, Cytokine concentrations in the serum shown as mean ± s.e.m. (two-tailed unpaired Student's *t* test, $*P < 0.05$, $**P < 0.01$, $***P < 0.001$, ns, not significant). All data are representative of two independent experiments.

Extended Data Table 1 | Data collection and refinement statistics for ALPK1-NTD structure (a) and Dali-search results of the structure (b)

a

	Se_ALPK1 (1-451)	Native_ALPK1 (1-446)
Data collection		
Space group	$P4_12_12$	$C2$
Cell dimensions		
a, b, c (Å)	185.52, 185.52, 334.11	169.26, 214.39, 173.15
α, β, γ (°)	90.00, 90.00, 90.00	90.00, 110.33, 90.00
Wavelength (Å)	0.97776	0.97853
Resolution (Å)[a]	50.00-4.46 (4.57-4.46)	50.00-2.59 (2.66-2.59)
R_{merge}	0.174 (1.033)	0.069 (0.660)
$I/\sigma(I)$	9.42 (2.32)	14.62 (1.99)
Completeness (%)	99.9 (100.0)	96.8 (95.8)
Redundancy	14.1 (13.0)	3.6 (3.6)
Refinement		
Resolution (Å)		48.89-2.59
No. of reflections		173,963
R_{work}/R_{free}		0.2420/0.2628
No. of atoms		
Protein		30,723
Water		0
B factors		
Protein		62.91
Water		0
r.m.s deviations		
Bond lengths (Å)		0.005
Bond angles (°)		0.599

b

No.	PDB chain	Z score	RMSD (Å)	Aligned residues	Identity (%)	Molecule name
1	5a6c-B	17.0	4.7	269	12	G-protein-signaling modulator 2, AFADIN
2	5a7d-B	16.9	6.4	269	12	PINS
3	5g05-O	16.4	4.8	271	10	Anaphase-promoting complex subunit 1
4	5dbk-A	15.6	3.5	239	13	Transcriptional regulator/TPR domain protein
5	3txm-A	14.9	4.2	237	11	26S proteasome regulatory complex subunit P42B
6	4ila-B	14.8	3.2	228	8	Response regulator aspartate phosphatase I
7	5o01-A	14.8	3.4	215	13	BKLC (Bacterial Kinesin-Light Chain-like)
8	4gyo-B	13.7	4.5	240	12	Response regulator aspartate phosphatase J
9	4ui9-F	13.4	4.0	233	8	Anaphase-promoting complex subunit 1
10	5l4k-Q	13.3	3.9	242	11	26S proteasome non-ATPase regulatory subunit 4

In **a**, single SeMet or native crystal was used for data collection and structure determination.
[a]Values in parentheses are for the highest-resolution shell.

LETTER

https://doi.org/10.1038/s41586-018-0465-8

Widespread intronic polyadenylation inactivates tumour suppressor genes in leukaemia

Shih-Han Lee[1,5], Irtisha Singh[2,3,5], Sarah Tisdale[1], Omar Abdel-Wahab[4], Christina S. Leslie[2] & Christine Mayr[1]*

DNA mutations are known cancer drivers. Here we investigated whether mRNA events that are upregulated in cancer can functionally mimic the outcome of genetic alterations. RNA sequencing or 3′-end sequencing techniques were applied to normal and malignant B cells from 59 patients with chronic lymphocytic leukaemia (CLL)[1–3]. We discovered widespread upregulation of truncated mRNAs and proteins in primary CLL cells that were not generated by genetic alterations but instead occurred by intronic polyadenylation. Truncated mRNAs caused by intronic polyadenylation were recurrent ($n = 330$) and predominantly affected genes with tumour-suppressive functions. The truncated proteins generated by intronic polyadenylation often lack the tumour-suppressive functions of the corresponding full-length proteins (such as DICER and FOXN3), and several even acted in an oncogenic manner (such as CARD11, MGA and CHST11). In CLL, the inactivation of tumour-suppressor genes by aberrant mRNA processing is substantially more prevalent than the functional loss of such genes through genetic events. We further identified new candidate tumour-suppressor genes that are inactivated by intronic polyadenylation in leukaemia and by truncating DNA mutations in solid tumours[4,5]. These genes are understudied in cancer, as their overall mutation rates are lower than those of well-known tumour-suppressor genes. Our findings show the need to go beyond genomic analyses in cancer diagnostics, as mRNA events that are silent at the DNA level are widespread contributors to cancer pathogenesis through the inactivation of tumour-suppressor genes.

In addition to DNA-based mutations, recent studies found that alterations in mRNA processing, including splicing, promote tumorigenesis[6]. In CLL, up to one-quarter of patients have mutations in ATM or SF3B1, but one-third have less than two mutated driver genes, and most patients (58%) only have a 13q deletion or have a normal karyotype[3,7–9]. Here, we investigated whether intronic polyadenylation (IPA) might serve as a new driver of tumorigenesis. Because 16% of genes in normal immune cells use IPA to generate truncated mRNAs that contribute to transcriptome diversity[2], we hypothesized that cancer-specific IPA would generate truncated proteins that lack essential domains, and thus, may phenocopy truncating (TR) mutations (Fig. 1a).

Using 3′-seq, a 3′-end sequencing method, on 44 samples including B cells from healthy donors and from patients with CLL, we identified 5,587 IPA isoforms, including 3,484 without previous annotation[1,2] (Extended Data Table 1 and Methods). We validated 4,630 IPA isoforms using RNA sequencing (RNA-seq) and additional 3′-seq data[2,10] (Extended Data Fig. 1a, b). To assess IPA usage in CLL, we first identified the normal B cell subset, the gene expression profile of which was most closely related to CLL cells. Lymphoid tissue-derived $CD5^+$ B cells were most similar (Extended Data Fig. 2), but clustered separately from CLL samples based on IPA site usage (Extended Data Fig. 1c). Using a generalized linear model (GLM), we identified 931 IPA events with significantly higher expression among 13 CLL samples, but low or absent expression in $CD5^+$ B cells[1,2] (Fig. 1b, Extended Data Fig. 1d). Because CLL IPAs are detectable by RNA-seq, we used an unrelated RNA-seq dataset to validate our CLL-IPA events[3] (Fig. 1c). We verified up to 71% of testable IPAs by this independent method and dataset (Extended Data Fig. 1d). For further analysis, we combined the datasets ($n = 59$ CLL samples) and focused only on CLL-IPAs that were present in more than 10% of the sample cohort resulting in 330 CLL-IPAs, derived from 306 genes (Fig. 1d, Supplementary Table 1). Although CLL-IPAs were detected in all CLL samples, one-third of the samples had a significantly higher number of CLL-IPAs (Fig. 1e, Extended Data Fig. 1e).

To investigate whether CLL-IPAs express truncated proteins, we performed western blots on 13 candidates. Whereas normal B cells only expressed the full-length proteins, the malignant B cells also expressed truncated proteins, the size of which was consistent with the predicted size of IPA-generated proteins (Fig. 2a, Extended Data Figs. 3 and 4).

To rule out that proteolytic cleavage truncates the proteins, we validated the presence of the IPA-generated truncated mRNAs (Extended Data Fig. 5a). Moreover, we were able to induce IPA isoform expression through the downregulation of splicing factors or through the inhibition of 5′ splice site recognition using an antisense oligonucleotide, indicating that deregulated mRNA processing can cause the expression of a truncated protein[11,12] (Extended Data Fig. 5b).

Many of the truncated proteins generated by CLL-IPAs are markedly similar to the predicted protein products produced by TR mutations, suggesting that CLL-IPAs may functionally mimic the outcome of genetic mutations (Fig. 2b, Extended Data Fig. 6a). To test this, we investigated the functional consequences of the expression of IPA and full-length protein isoforms of four candidates in malignant B cells. CARD11 is a positive regulator of the NF-κB pathway and is important for lymphocyte survival and proliferation[13]. We observed substantial CARD11 IPA protein production, compared to only slightly increased CARD11 IPA mRNA expression, indicating that the truncated protein is more stable and may activate the NF-κB signalling pathway more potently than the full-length protein[14] (Fig. 2a). To test this, we exclusively knocked down either full-length or CARD11 IPA in a malignant B cell line that expresses CARD11 IPA at comparable levels to those expressed by CLL cells (Extended Data Fig. 6b, c). We measured phosphorylated p65 (also known as RELA) to assess NF-κB activity and found significantly lower activity after knockdown of CARD11 IPA than of the full-length protein (Fig. 2c, Extended Data Fig. 6d). Thus, CARD11 IPA activates NF-κB more potently than full-length CARD11, suggesting that it may mimic activating mutations present in high-grade lymphomas[13]. CARD11 IPA may contribute to NF-κB activation in CLL, in which the signalling components are rarely mutated[15].

DICER IPA generates a truncated protein that partially lacks the RNase IIIB domain responsible for microRNA (miRNA) processing[16] (Fig. 2b). In contrast to full-length DICER, DICER IPA entirely lacks miRNA cleavage ability and mimicked TR mutations that remove both RNase III domains[16] (Fig. 2b, d, Extended Data Fig. 6e, f). Although DICER IPA does not act in a dominant-negative manner, its expression reduces functional DICER protein, thus potentially decreasing endogenous miRNA expression.

The tumour-suppressor gene (TSG) *MGA* is targeted by TR mutations in CLL and solid cancers[3,7,17] (Fig. 2b). MGA negatively regulates

[1]Cancer Biology and Genetics Program, Memorial Sloan Kettering Cancer Center, New York, NY, USA. [2]Computational and Systems Biology Program, Memorial Sloan Kettering Cancer Center, New York, NY, USA. [3]Tri-I Program in Computational Biology and Medicine, Weill Cornell Graduate College, New York, NY, USA. [4]Human Oncology and Pathogenesis Program, Memorial Sloan Kettering Cancer Center, New York, NY, USA. [5]These authors contributed equally: Shih-Han Lee, Irtisha Singh. *e-mail: mayrc@mskcc.org

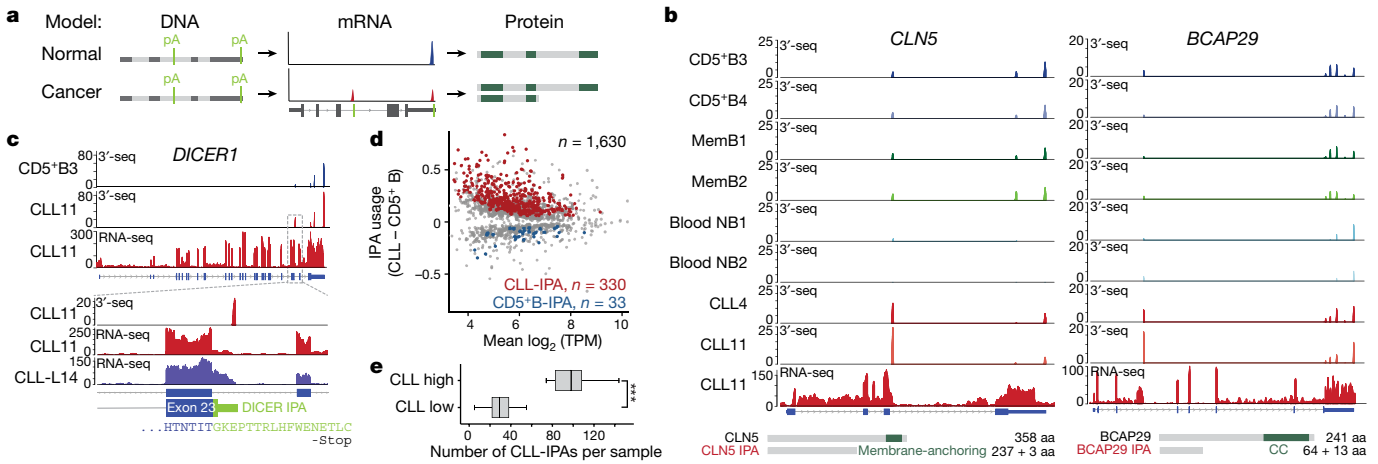


Fig. 1 | Hundreds of genes generate recurrent CLL-IPAs. a, Schematic showing full-length mRNA and protein expression in normal cells and the generation of a truncated mRNA and protein through cancer-specific IPA, despite no difference in DNA sequence. Polyadenylation sites (pA) are shown in light green. Loss of essential protein domains (dark green boxes) through cancer-gained IPA may inactivate TSGs, thus contributing to cancer pathogenesis. **b**, Representative CLL-IPAs (from $n = 330$) are shown. mRNA 3′ ends detected by 3′-seq are depicted as peaks, the heights of which correspond to transcript abundance shown in transcripts per million (TPM). The bottom panel shows RNA-seq reads and numbers correspond to read counts. Full-length and IPA-generated truncated proteins are depicted in grey, known domains are shown in green and the domains lost through IPA are named. For CLL-IPA, the number of retained and novel amino acids (aa) and amino acids of full-length proteins are given. CC, coil–coil; MemB, memory B cells, NB, naive B cells. **c**, Representative RNA-seq tracks from two independent CLL datasets are shown as in **b**; one is indicated by 'L' before the patient number (CLL-L14). B3 denotes donor 3. Zoomed-in view shows the exonized part of intron 23 of *DICER1* (green). **d**, Difference in relative abundance (usage) of IPA isoforms between CLL and normal $CD5^+$ B cells. A GLM was used to identify significant events. CLL-IPAs with significantly higher usage are shown in red (false discovery rate (FDR)-adjusted $P < 0.1$, usage difference ≥ 0.05, TPM in $CD5^+$ B < 8) and $CD5^+$ B-IPAs are shown in blue. Grey denotes IPAs present in CLL and $CD5^+$ B cells without significantly different usage. **e**, Number of CLL-IPAs per sample is shown as box plots, in which the horizontal line denotes the median; boxes denote the 25th and 75th percentiles; error bars denote the range. CLL high, $n = 21/59$, median of CLL-IPAs/sample = 98 versus CLL low, $n = 38/59$, median = 29. $***P = 6 \times 10^{-10}$, two-sided Mann–Whitney *U*-test.

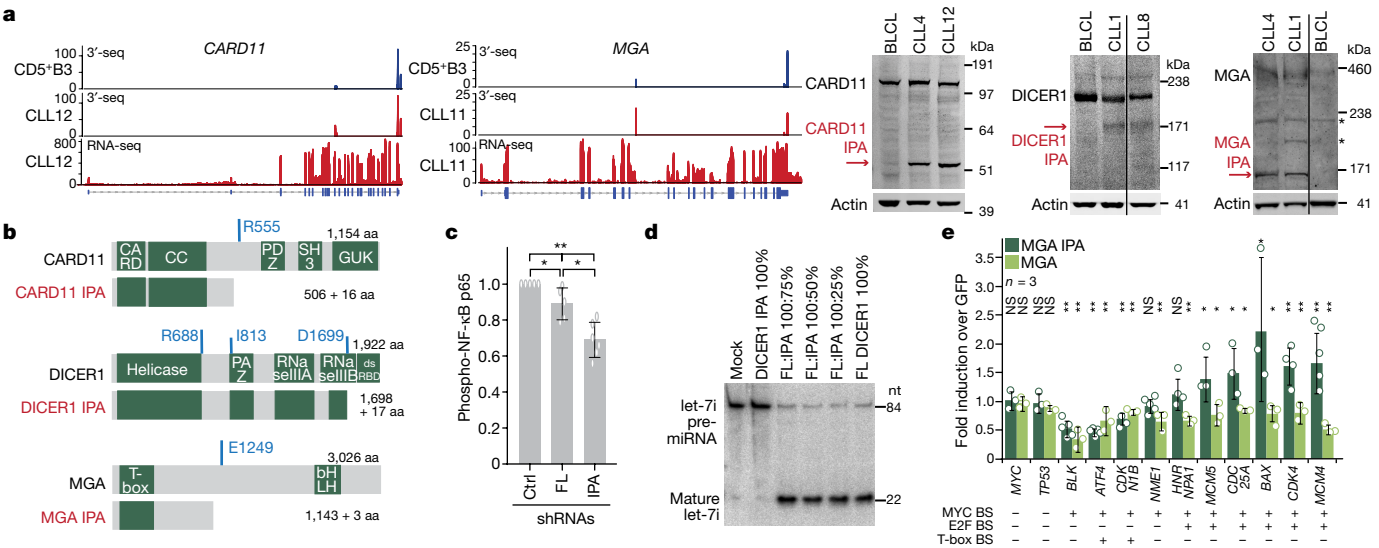


Fig. 2 | IPA-generated truncated proteins resemble the protein products of truncating DNA mutations and have cancer-promoting properties. **a**, RNA-seq and 3′-seq data of functionally validated CLL-IPAs ($n = 5$) as in Fig. 1b. The remaining tracks are shown in Extended Data Fig. 3. Endogenous full-length proteins are detected by western blot analysis in CLL and normal B cells (B lymphoblastoid cells; BLCL), whereas IPA-generated truncated proteins (red arrows) are only present in primary CLL cells. Actin was used as loading control. The experiment was replicated with similar results (*CARD11*, $n = 4$, *DICER*, $n = 3$, *MGA*, $n = 2$). For gel source data see Supplementary Fig. 1. Asterisks denote an unspecific band. **b**, Protein models are shown as in Fig. 1b. The amino acid positions of recurrent TR mutations are shown in blue. **c**, Endogenous phospho-NF-κB p65 levels are shown as normalized mean fluorescent intensity (MFI) values after short hairpin RNA (shRNA)-mediated knockdown of full-length (FL) CARD11 and CARD11 IPA. $n = 5$ (FL shRNA1 and control (ctrl) shRNA) or $n = 6$ (IPA; shRNA2 $n = 3$, shRNA3 $n = 3$) biologically independent experiments. Data are mean ± s.d. $**P = 0.002$, two-sided Kruskal–Wallis test; *P* value of two-sided Mann–Whitney *U*-test was adjusted for multiple testing, *adjusted $P = 0.036$. **d**, miRNA cleavage assay, performed twice with similar results, showing processing of pre-let-7i into mature let-7i by V5-DICER. Mock indicates that no protein was added. V5-DICER IPA shows a complete loss of function, but no dominant-negative activity. nt, nucleotides. **e**, qRT–PCR of endogenous MYC target genes after expression of full-length or MGA IPA in Raji cells. Shown are *GAPDH*-normalized values as mean ± s.d. from three biological replicates, each performed in technical triplicates. $*P < 0.05$, $**P < 0.001$, two-sided *t*-test for independent samples. NS, not significant. Exact *P* values are shown in Supplementary Fig. 1. MGA represses all MYC target genes. BS, binding sites.

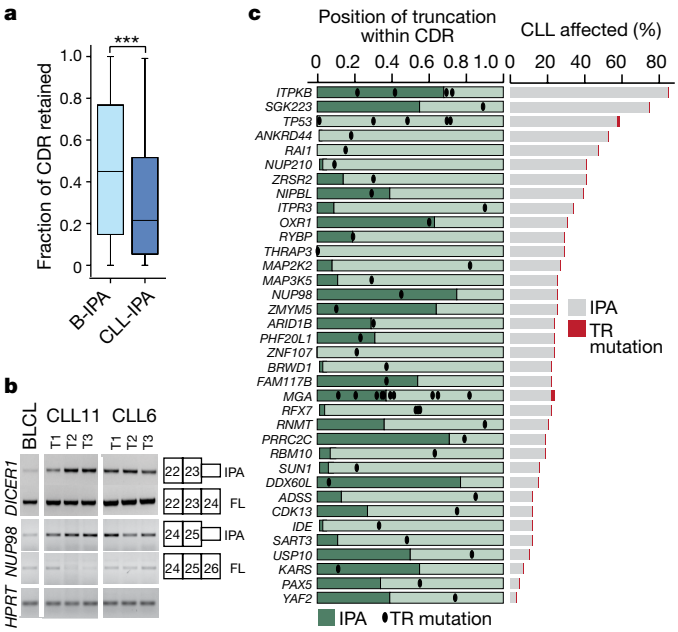


Fig. 3 | TSGs are enriched among CLL-IPAs. CLL-IPAs and TR mutations in CLL target the same genes but in different patients. **a**, The fraction of the retained coding region (CDR) is shown for genes that generate CLL-IPAs ($n = 306$, median fraction of retained CDR = 0.21; 112 amino acids) and B-IPAs ($n = 2{,}690$, median fraction of retained CDR = 0.45; 221 amino acids). ***$P = 1 \times 10^{-16}$, two-sided Mann–Whitney *U*-test. Box plots are as in Fig. 1e. **b**, RT–PCR analysis showing the expression of full-length and IPA isoforms for two TSGs (*DICER1* and *NUP98*) in samples from two patients with CLL that were collected over several years. CLL11: T1, 17 months after diagnosis, T2, 24 months, T3, 44 months; CLL6: T1, 16 months, T2, 49 months, T3, 91 months (42 months after treatment). Shown are the exons that contain primers for amplifications of the products. BLCL serve as control cells. The expression of *HPRT* was used as a loading control. **c**, Genes that are targeted by TR mutations in CLL and CLL-IPAs are shown ($n = 36$). Dark green bars indicate the fraction of retained CDRs for each IPA-generated protein. Black dots indicate the positions of TR mutations in CLL. CLL-IPAs occur mostly in the vicinity of or upstream of TR mutations. $P = 0.004$, two-sided Wilcoxon rank-sum test. Right, the fraction of CLL samples affected is shown for each gene and represents the fraction of CLL samples (out of 59) with significantly upregulated expression of the IPA isoform (CLL-IPA, grey; TR mutations, red).

the MYC transcriptional program and represses genes with MYC- and E2F-binding sites in a Polycomb-dependent manner[18,19]. Expression of MGA from constructs validated MGA IPA detected in CLL cells and confirmed the repressive effect of MGA on MYC target gene expression in malignant B cells (Fig. 2e, Extended Data Fig. 6g). Notably, on genes with binding sites for both MYC and E2F, MGA IPA acts as dominant-negative regulator of full-length MGA as it significantly induced the expression of 5 out of 6 genes in cells that endogenously express full-length MGA (Fig. 2e). However, as MGA IPA retains the N-terminal T-box, it still acts as a repressor on T-box target genes (Fig. 2e).

Lastly, the IPA isoform of the transcriptional repressor FOXN3[20] derepressed its oncogenic targets *MYC* and *PIM2* (Extended Data Figs. 3, 6h–j). In summary, the CLL-IPA-generated proteins can contribute to cancer pathogenesis in various ways. Their generation can reduce the expression of functional TSGs (DICER and FOXN3 IPA) or they behave as dominant-negatives, thus acting in an oncogenic manner (MGA IPA).

Because all functionally validated CLL-IPAs produced dysfunctional proteins, we investigated whether this is a general feature. We compared the retained fraction of amino acids of IPA isoforms present in normal B cells (B-IPA, $n = 2{,}690$) with CLL-IPAs. Although the protein size of full-length proteins targeted by IPA was similar, CLL-IPAs lose significantly more amino acids than B-IPAs (Fig. 3a, Extended Data Fig. 7a). This suggests that IPA in normal cells contributes to proteome diversity[2], whereas CLL-IPAs tend to produce dysfunctional proteins.

Because genes targeted by TR mutations are often TSGs[5] (Extended Data Fig. 7b), we investigated whether TSGs are overrepresented among CLL-IPAs. Compared to control groups with matched protein sizes, there was a significant enrichment of TSGs among CLL-IPAs ($P = 3 \times 10^{-5}$; Extended Data Fig. 7c–f). Importantly, IPA-generated truncated proteins usually lack either more or a comparable number of amino acids compared to truncated proteins generated by TR mutations, suggesting the IPA isoforms are probably inactive (Extended Data Fig. 7c). However, for CLL-IPAs to inactivate TSGs, they must also be stably expressed. For 11 out of 12 tested CLL-IPAs, we observed stable expression at the mRNA or protein level over a four-year time span (Fig. 3b, Extended Data Fig. 5c, d), indicating that they have the potential to inactivate TSGs.

In addition to TSGs in general, we found that genes inactivated by TR mutations in CLL are enriched among CLL-IPAs[3,7,8] (Fig. 3c, Extended Data Fig. 7g). Notably, the fraction of samples affected by CLL-IPA was substantially larger than the number of CLL samples affected by TR mutations (3.0–85% versus 0.13–2.0%; Fig. 3c, right). This indicates that TR mutations and CLL-IPAs target the same genes in different patient groups, thus substantially expanding the proportion of patients with protein truncations in potential drivers.

To rule out the possibility that CLL-IPAs are caused by somatic mutations, we examined the presence of DNA mutations in the CLL-IPA genes. Two genes were targeted by TR mutations and IPA in the same patient. Notably, the two inactivation mechanisms are predicted to generate different truncated protein products, suggesting that they occurred independently[3] (Extended Data Fig. 7h, i). The mutation data also enabled us to associate CLL-IPAs with specific somatic mutations. CLL samples with a high number of IPA were enriched in SF3B1 mutations, but they were independent of IGVH mutational status (Extended Data Fig. 7j–l).

Because of the enrichment of known TSGs among CLL-IPAs, we examined whether CLL-IPAs may enable us to identify novel TSGs. We selected CLL-IPAs present in at least 20% of CLL samples ($n = 199$, generated from 190 genes; Fig. 4a, Supplementary Tables 1 and 2). We next investigated whether these genes are inactivated by TR mutations in solid cancers using mutations from more than 86,000 tumours, compiled by the MSK cbio portal[4]. We observed that 72% of these genes are frequently affected by TR mutations in solid tumours and call them novel TSG candidates (136 out of 190; Fig. 4b). This is a significant

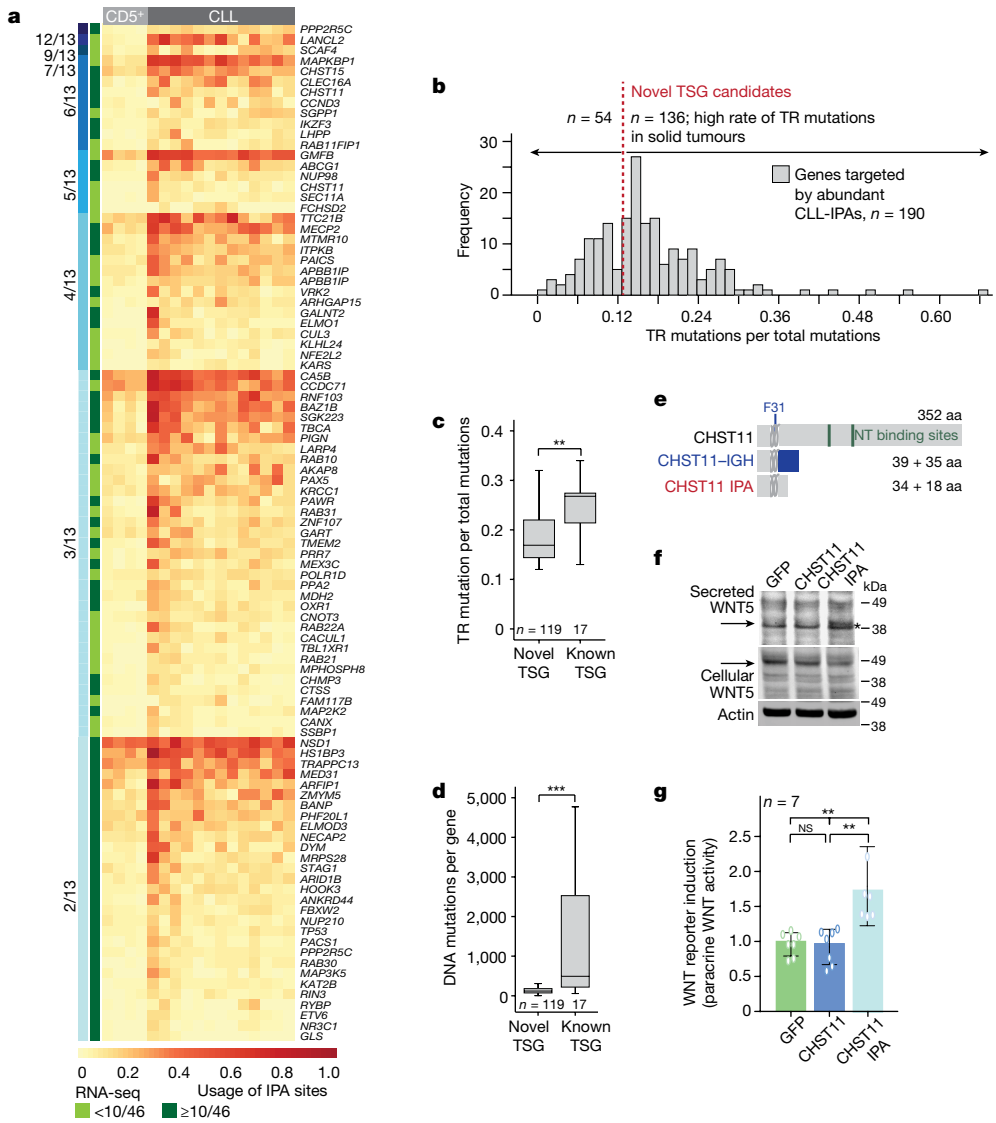


Fig. 4 | Novel TSG candidates are inactivated in CLL at the mRNA level and in solid tumours at the DNA level. a, Colour-coded IPA usage for a subset of CLL-IPAs (97 out of 199 of samples with significant expression of IPA in ≥20% of CLL samples). Gene names and number of affected CLL samples per CLL-IPA are indicated (blue bars, 3′-seq, green bars, RNA-seq). **b**, Truncating mutation rates (number of TR mutations per total mutations) in solid tumours, obtained from the MSK cbio portal for genes that generate abundant CLL-IPAs, partially shown in **a**. The bimodal distribution was separated at the local minimum (TR mutation per total mutations = 0.12, red line) into two gene groups: those rarely targeted by TR mutations and those with high TR mutation rates in solid cancers, defined as novel TSG candidates. **c**, TR mutation rates of known and novel TSG candidates. $^{**}P = 0.0002$, two-sided Mann–Whitney *U*-test. Box plots are as in Fig. 1e. **d**, As shown in **c**, but for overall mutation rates. $^{***}P = 1 \times 10^{-10}$, two-sided Mann–Whitney *U*-test. **e**, CHST11 protein models, shown as in Fig. 2b. Loops depict membrane domains. A chromosomal translocation in CLL results in fusion of the immunoglobulin heavy chain locus (IGH) with a truncated CHST11[23]. NT, nucleotide. **f**, Western blot of WNT5B, performed once, shown as in Fig. 2a, from cell lysates of or conditioned medium obtained from B cells stably expressing green fluorescent protein (GFP), GFP-tagged CHST11 or GFP-tagged CHST11 IPA. Conditioned medium from cells expressing CHST11 IPA contains unglycosylated WNT5B[25]. Asterisk, unspecific band. **g**, Conditioned medium from samples described in **f** was added to HEK293T cells expressing a WNT reporter, and normalized luciferase activity is shown. Data are mean ± s.d. from $n = 7$ biologically independent experiments. $^{**}P = 0.002$, two-sided Kruskal–Wallis test; *P* value of two-sided Mann–Whitney *U*-test was adjusted for multiple testing, **adjusted $P = 0.002$.

enrichment over background and the list contains 17 known TSGs and 119 novel TSG candidates[5] (Extended Data Fig. 8a, b). Again, CLL-IPAs lack more or a comparable number of amino acids as the proteins produced by TR mutations, suggesting that CLL-IPAs inactivate the functions of these genes (Extended Data Fig. 8a).

Although the TR mutation rates of the novel TSG candidates were comparable with known TSGs found at the lower end of the spectrum, their protein size and overall mutation rates were substantially lower (Fig. 4c, d, Extended Data Fig. 8c). This may explain why these potentially cancer-relevant genes have been overlooked thus far[21]. As they are targeted at the mRNA level in leukaemia and at the DNA level in solid cancers, they should be considered as a novel class of TSG candidates. To support this, we functionally validated a highly recurrent CLL-IPA isoform that affected a poorly known cancer gene. *CHST11* encodes a Golgi-associated carbohydrate sulfotransferase that modifies chondroitin on the surface of WNT-expressing cells. The modification results in the binding of secreted WNT and prevents its paracrine action[22]. CHST11 IPA lacks catalytic activity, but retains the cytoplasmic tail[23] (Fig. 4e, Extended Data Fig. 8d). As exclusive expression of the cytoplasmic tail of Golgi enzymes inhibited localization of full-length enzymes[24], we hypothesized that CHST11 IPA may act in a dominant-negative manner. We expressed CHST11 and CHST11 IPA, collected the conditioned media, and detected secreted WNT in medium only after expressing CHST11 IPA[25] (Fig. 4f, Extended Data Fig. 8e, f). The

conditioned medium activated a WNT reporter in HEK293T cells (Fig. 4g), demonstrating that CHST11 IPA enabled paracrine WNT action on neighbouring cells through dominant-negative action. Thus, in addition to mutations in the WNT pathway[26], CLL-IPAs may also contribute to WNT activation in CLL.

A member of this new class of TSGs was recently found in breast cancers, in which tumour-specific expression of MAGI3 IPA generates a truncated protein with dominant-negative activity[27] (Extended Data Fig. 9a). Combined with our findings on T-lineage acute lymphoblastic leukaemia (T-ALL), in which we detected more than 100 IPA isoforms (Extended Data Fig. 9b), these data indicate that cancer-upregulated IPA isoforms are not restricted to CLL.

In summary, we found that TSGs can be inactivated, either fully or partially, by IPA. Even partial loss of TSG function was shown to contribute crucially to tumorigenesis[28]. As CLL-IPAs are not generated by DNA mutations in their corresponding transcription units, DNA and mRNA alterations occur in different patient groups. In CLL, the fraction of patients with TSGs that are inactivated by CLL-IPAs is considerably larger than those with TSGs disrupted by TR mutations (Fig. 3c); thus, CLL-IPAs substantially expand the number of patients with affected drivers. Moreover, these data identify a class of TSGs that is predominantly inactivated at the mRNA rather than the DNA level[27]. Thus, our study demonstrates that cancer-gained changes in mRNA processing can functionally mimic the effects of somatic mutations and shows the need to go beyond genomic analyses in cancer diagnostics.

Online content

Any Methods, including any statements of data availability and Nature Research reporting summaries, along with any additional references and Source Data files, are available in the online version of the paper at https://doi.org/10.1038/s41586-018-0465-8.

Received: 18 July 2017; Accepted: 17 July 2018;
Published online 27 August 2018.

1. Lianoglou, S., Garg, V., Yang, J. L., Leslie, C. S. & Mayr, C. Ubiquitously transcribed genes use alternative polyadenylation to achieve tissue-specific expression. *Genes Dev.* **27**, 2380–2396 (2013).
2. Singh, I. et al. Widespread intronic polyadenylation diversifies immune cell transcriptomes. *Nat. Commun.* **9**, 1716 (2018).
3. Landau, D. A. et al. Mutations driving CLL and their evolution in progression and relapse. *Nature* **526**, 525–530 (2015).
4. Cerami, E. et al. The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. *Cancer Discov.* **2**, 401–404 (2012).
5. Davoli, T. et al. Cumulative haploinsufficiency and triplosensitivity drive aneuploidy patterns and shape the cancer genome. *Cell* **155**, 948–962 (2013).
6. Dvinge, H., Kim, E., Abdel-Wahab, O. & Bradley, R. K. RNA splicing factors as oncoproteins and tumour suppressors. *Nat. Rev. Cancer* **16**, 413–430 (2016).
7. Puente, X. S. et al. Whole-genome sequencing identifies recurrent mutations in chronic lymphocytic leukaemia. *Nature* **475**, 101–105 (2011).
8. Quesada, V. et al. Exome sequencing identifies recurrent mutations of the splicing factor SF3B1 gene in chronic lymphocytic leukemia. *Nat. Genet.* **44**, 47–52 (2011).
9. Stilgenbauer, S., Bullinger, L., Lichter, P., Döhner, H. & the German CLL Study Group (GCLLSG). Genetics of chronic lymphocytic leukemia: genomic aberrations and V(H) gene mutation status in pathogenesis and clinical course. *Leukemia* **16**, 993–1007 (2002).
10. Gruber, A. J. et al. A comprehensive analysis of 3′ end sequencing datasets reveals novel polyadenylation signals and the repressive role of heterogeneous ribonucleoprotein C on cleavage and polyadenylation. *Genome Res.* **26**, 1145–1159 (2016).
11. Vorlová, S. et al. Induction of antagonistic soluble decoy receptor tyrosine kinases by intronic polyA activation. *Mol. Cell* **43**, 927–939 (2011).
12. Zarnack, K. et al. Direct competition between hnRNP C and U2AF65 protects the transcriptome from the exonization of Alu elements. *Cell* **152**, 453–466 (2013).
13. Lenz, G. et al. Oncogenic CARD11 mutations in human diffuse large B cell lymphoma. *Science* **319**, 1676–1679 (2008).
14. Bertin, J. et al. CARD11 and CARD14 are novel caspase recruitment domain (CARD)/membrane-associated guanylate kinase (MAGUK) family members that interact with BCL10 and activate NF-κB. *J. Biol. Chem.* **276**, 11877–11882 (2001).
15. Mansouri, L., Papakonstantinou, N., Ntoufa, S., Stamatopoulos, K. & Rosenquist, R. NF-κB activation in chronic lymphocytic leukemia: a point of convergence of external triggers and intrinsic lesions. *Semin. Cancer Biol.* **39**, 40–48 (2016).
16. Rakheja, D. et al. Somatic mutations in DROSHA and DICER1 impair microRNA biogenesis through distinct mechanisms in Wilms tumours. *Nat. Commun.* **5**, 4802 (2014).
17. De Paoli, L. et al. MGA, a suppressor of MYC, is recurrently inactivated in high risk chronic lymphocytic leukemia. *Leuk. Lymphoma* **54**, 1087–1090 (2013).
18. Hurlin, P. J., Steingrìmsson, E., Copeland, N. G., Jenkins, N. A. & Eisenman, R. N. Mga, a dual-specificity transcription factor that interacts with Max and contains a T-domain DNA-binding motif. *EMBO J.* **18**, 7019–7028 (1999).
19. Ogawa, H., Ishiguro, K., Gaubatz, S., Livingston, D. M. & Nakatani, Y. A complex with chromatin modifiers that occupies E2F- and Myc-responsive genes in G0 cells. *Science* **296**, 1132–1136 (2002).
20. Huot, G. et al. CHES1/FOXN3 regulates cell proliferation by repressing PIM2 and protein biosynthesis. *Mol. Biol. Cell* **25**, 554–565 (2014).
21. Vogelstein, B. et al. Cancer genome landscapes. *Science* **339**, 1546–1558 (2013).
22. Nadanaka, S., Kinouchi, H., Taniguchi-Morita, K., Tamura, J. & Kitagawa, H. Down-regulation of chondroitin 4-*O*-sulfotransferase-1 by Wnt signaling triggers diffusion of Wnt-3a. *J. Biol. Chem.* **286**, 4199–4208 (2011).
23. Schmidt, H. H. et al. Deregulation of the *carbohydrate* (*chondroitin 4*) *sulfotransferase 11* (*CHST11*) gene in a B-cell chronic lymphocytic leukemia with a t(12;14)(q23;q32). *Oncogene* **23**, 6991–6996 (2004).
24. Milland, J., Russell, S. M., Dodson, H. C., McKenzie, I. F. & Sandrin, M. S. The cytoplasmic tail of α1,3-galactosyltransferase inhibits Golgi localization of the full-length enzyme. *J. Biol. Chem.* **277**, 10374–10378 (2002).
25. Kessenbrock, K. et al. A role for matrix metalloproteinases in regulating mammary stem cell function via the Wnt signaling pathway. *Cell Stem Cell* **13**, 300–313 (2013).
26. Wang, L. et al. Somatic mutation as a mechanism of Wnt/β-catenin pathway activation in CLL. *Blood* **124**, 1089–1098 (2014).
27. Ni, T. K. & Kuperwasser, C. Premature polyadenylation of MAGI3 produces a dominantly-acting oncogene in human breast cancer. *eLife* **5**, e14730 (2016).
28. Berger, A. H., Knudson, A. G. & Pandolfi, P. P. A continuum model for tumour suppression. *Nature* **476**, 163–169 (2011).

Acknowledgements This work was funded by the NCI grant U01-CA164190 (to C.M. and C.S.L.), a Starr Cancer Foundation grant (to C.M. and C.S.L.), the Innovator Award of the Damon Runyon-Rachleff Cancer Foundation and the Island Outreach Foundation (DRR-24-13; to C.M.), the NIH Director's Pioneer Award (DP1-GM123454, to C.M.), the Pershing Square Sohn Cancer Research Alliance (to C.M.) and the MSK Core Grant (P30 CA008748). We are grateful to V. K. Modi for access to lymphatic tissue, to D. A. Landau for providing CLL RNA-seq data and sample identities to validate our findings, and to C. Wu and D. Neuberg for clinical outcome analyses. We thank J. Mendell and V. Narry Kim for providing the V5-DICER construct and the DICER knockout cells, J. Chaudhuri for critical reading of the manuscript, and the members of the Mayr laboratory for discussions.

Reviewer information *Nature* thanks M. Muschen and the other anonymous reviewer(s) for their contribution to the peer review of this work.

Author contributions S.-H.L. organized and collected the samples and generated the libraries. S.-H.L. and S.T. performed and analysed all validation and functional experiments and contributed to study design. I.S. performed all of the computational analyses with respect to identification of IPA isoforms and their integration with published CLL datasets with input from C.S.L. and C.M. O.A.-W. provided the CLL samples and some of the CLL RNA-seq data. C.M. conceived the study and integrated CLL-IPAs with mutation analysis of solid cancers. C.M., S.-H.L. and S.T. wrote the manuscript with input from all authors.

Competing interests The authors declare no competing interests.

Additional information
Extended data is available for this paper at https://doi.org/10.1038/s41586-018-0465-8.
Supplementary information is available for this paper at https://doi.org/10.1038/s41586-018-0465-8.
Reprints and permissions information is available at http://www.nature.com/reprints.
Correspondence and requests for materials should be addressed to C.M.
Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

METHODS

No statistical methods were used to predetermine sample size. The experiments were not randomized, and investigators were not blinded to allocation during experiments and outcome assessment.

Samples for 3′-seq and RNA-seq analyses. Samples were obtained from untreated patients with CLL seen at Memorial Sloan Kettering Cancer Center, New York (Extended Data Table 1a). All patients provided written informed consent before participating in the study. The sample collection was approved by the Institutional Review Board of Memorial Sloan Kettering Cancer Center. Peripheral blood mononuclear cells from CLL samples with a minimum white blood cell count of 75,000 per microlitre were isolated by Ficoll (GE Healthcare) gradient centrifugation at 400 r.c.f. for 30 min, followed by two washes in PBS at room temperature. Cells were treated with red blood cell lysis buffer (155 mM NH_4Cl, 12 mM $NaHCO_3$, 0.1 mM EDTA) for 5 min at room temperature and were washed twice with PBS. Pure CLL B cells were obtained from peripheral blood mononuclear cells using B-CLL isolation kit (Miltenyi Biotec). This selected untouched CLL cells using a cocktail of magnetic beads coated with CD2, CD3, CD4, CD14, CD15, CD16, CD56, CD61, CD235a, FcεRI and CD34. The purity of CLL B cells ($CD5^+$ and $CD19^+$) was analysed by FACS and the cells were immediately dissolved in TRI Reagent (Ambion) for RNA extraction, followed by 3′-seq or RNA-seq library preparation.

For longitudinal analyses, samples from two patients were investigated at different time points during the course of the disease. CLL11, time point 1 (T1) 17 months after diagnosis, T2, 24 months after diagnosis, T3, 44 months after diagnosis. The patient was not treated with chemotherapy during the sample collection period. CLL6: T1, 16 months after diagnosis, T2, 49 months, T3, 91 months (42 months after chemotherapeutic treatment).

In addition to the newly generated CLL 3′-seq data, we also used 3′-seq data from normal tissues, cell lines and immune cell subsets that we generated previously[1,2] (Extended Data Table 1b).

We performed RNA-seq on 11 CLL samples (Extended Data Table 1a) and obtained access to a previously published RNA-seq dataset from 44 patients with CLL[3] that was provided by D. A. Landau. RNA-seq data from normal immune cells were obtained from samples we generated previously[2] (Extended Data Table 1c). For validation of 3′-seq data, we also used publicly available RNA-seq (tonsil-derived NB, GSE45982 (GSM1129340–GSM1129347)[29], blood-derived NB, ERR431624, ERR431586[30], $CD3^+$ T cells, GSM1576415[31]) and 3′-seq data[10].

For RNA-seq-based identification of IPA isoforms expressed in T-ALL, we used publicly available RNA-seq data from 10 primary T-ALL samples and 2 whole human thymus extracts (GSE57982)[32].

FACS sorting of immune cell populations. Cells were washed with ice-cold PBS once, incubated with appropriate fluorochrome-conjugated antibodies for 30 min at 4 °C and washed twice with ice-cold PBS containing 0.5% FCS. The following antibodies were used: anti-CD3-PE (mouse, BD Biosciences, 555333), anti-CD5-FITC (mouse, BD Biosciences, 555352), anti-CD14-PECy7 (mouse, ebioscience, 25-0149-42), anti-CD19-APC (mouse, BD Biosciences, 555415), anti-CD27-PE (mouse BD Biosciences, 555441), anti-CD38-APC (mouse, BD Biosciences, 555462), anti-CD38-FITC (mouse, BD Biosciences, 555459). Surface protein expression was detected by a BD FACSCalibur cell analyser (BD Biosciences) and data were analysed using the FlowJo software.

3′-seq and RNA-seq analyses. 3′-seq libraries were generated as previously described and sequenced with Illumina HiSeq using single-end 50-nucleotide reads[1,2]. RNA-seq libraries were prepared at the Weill Cornell and the MSKCC Genomics core facilities.

Analysis of 3′-seq data was performed as described previously[1] with a few modifications that have been extensively described[2]. In brief, a gene is considered to be expressed if either the IPA isoform (≥5 TPM) or the full-length isoform (≥5.5 TPM) were expressed in 75% of the samples of a particular cell type. We focused our analysis on robustly expressed transcript isoforms and filtered 3′-seq peaks according to their usage. Robustly expressed 3′UTR isoforms that are part of the atlas are expressed with at least 3 TPM in at least one sample and each peak combines at least 10% of all reads that map to the 3′UTR. Robustly expressed IPA isoforms that are part of the atlas are expressed with 5 TPM or more and had ≥0.1 IPA site usage in at least one sample. IPA site usage is the relative expression of each IPA isoform with respect to the total expression of 3′UTR isoforms (all reads that fall into robust 3′UTR peaks are summed up). We only analysed IPA isoforms of protein coding genes.

Validation of IPA isoforms using external data sources. To obtain evidence of IPA isoforms from independent methods, we first used RNA-seq data obtained from the same RNA or from the same cell type to identify IPA isoforms. We used the coordinates of the IPA events obtained from 3′-seq and tested the RNA-seq read counts in windows of 100 nucleotides located upstream and downstream of the IPA peak using a GLM[2] (Extended Data Fig. 1a). The windows were separated by 51 nucleotides centred on the first nucleotide of the polyadenylation signal. Not all IPA isoforms could be tested. For example, if the defined windows overlapped with an annotated exon, the IPA event was excluded from further analysis. An IPA isoform was considered present if we detected a significant difference in read counts within the upstream and downstream windows (adjusted $P < 0.1$) using DESeq. This analysis was also used to validate CLL-gained IPA events in an independent CLL dataset.

We further regarded an IPA isoform as validated if reads that overlap with IPA peaks had at least four untemplated adenosines in the RNA-seq data and a polyadenylation signal (or one of its variants)[33] was detected within 50 nucleotides upstream of the read. In addition, we considered IPA isoforms as validated if we detected read evidence in independent 3′-seq datasets[10]. As no previous 3′-seq data exist for many of our cell types, we also included highly expressed (≥10 TPM and ≥0.1 IPA site usage) IPA isoforms with an upstream polyadenylation signal (AAUAAA and its variants)[33] in our downstream analysis.

Identification of the normal counterpart of CLL and of CLL-IPAs. Hierarchical clustering was performed on the normal human B cell subsets derived from lymphoid tissues or peripheral blood and CLL samples using RNA-seq derived mRNA expression levels (quantile normalized $\log_2$ reads per kilobase of transcript per million mapped reads (RPKM)). Genes expressed with greater than 5.5 RPKM in 75% of normal B cells or any of the CLL samples went into the analysis. The 20% most variable genes by median absolute deviation across the dataset were used for the clustering. The heat map was generated using aheatmap (http://cran.r-project.org/package=NMF) with row scaling. This analysis showed that lymphoid-tissue derived $CD5^+$ B cells are most closely related in their gene expression profile to CLL cells (Extended Data Fig. 2).

We performed hierarchical unsupervised clustering of CLL and control samples based on IPA site usage to test whether IPA site usage separates normal and malignant B cells (Extended Data Fig. 1c). The top 20% most variable genes by median absolute deviation across all the $CD5^+$ B and CLL samples were used. This analysis showed two main clusters: Four CLL samples (CLL4, CLL7, CLL11 and CLL12) clustered separately from the rest of the samples. However, within the rest of the samples, the control group ($CD5^+$ B) clustered separately. The four CLL samples that differed the most from the rest of the samples had a high number of significantly upregulated IPA isoforms (CLL high: median number of CLL-IPAs per sample, $n = 100$; range, $n = 42$–274), whereas the remaining samples had a low number of CLL-IPAs (CLL low: median, $n = 9$; range, $n = 5$–28; Extended Data Fig. 1e).

To identify CLL-upregulated IPA isoforms, we applied a GLM[1,2,34] and tested usage of each IPA isoform between the normal B cell group and each CLL sample. We only considered IPA isoforms that were significantly upregulated in CLL (FDR-adjusted $P < 0.1$, usage difference between CLL and $CD5^+$ B ≥ 0.05) and were either not or lowly expressed in $CD5^+$ B cells (TPM < 8, corresponding to 75% quantile for $CD5^+$ B TPM). This resulted in 931 significantly upregulated IPA events observed in 13 CLL samples. $n = 454$ IPA events were detected in only a single sample and were regarded as non-recurrent, whereas 477 IPA events occurred in more than one sample (≥2 out of 13), and were considered recurrent events by 3′-seq (Extended Data Fig. 1d). The recurrent events resulted in 168 recurrent CLL-IPA isoforms.

As CLL-IPAs are detectable by RNA-seq, we used an independent RNA-seq dataset containing 46 CLL samples for validation[3]. We verified up to 71% of testable IPAs by this independent method and dataset. Because of the high validation rate, we combined the two datasets ($n = 59$ CLL samples) and focused on CLL-IPAs present in more than 10% of the whole CLL sample cohort. This resulted in 330 CLL-upregulated IPA isoforms, derived from 306 genes (Supplementary Table 1). The list of 330 CLL upregulated IPA isoforms contains the 168 CLL-IPAs identified in at least 2 out of 13 3′-seq samples, but contains also CLL-IPA isoforms detected in one 3′-seq and in at least five additional RNA-seq samples (≥6 out of 59 total samples).

We detected 33 IPA events that showed significantly higher IPA site usage in $CD5^+$ B cells compared with CLL. IPA site usage was required to be higher than in 2 CLL samples (TPM <10, corresponding to 75% quantile for CLL TPM; FDR-adjusted $P < 0.1$, usage difference between CLL and $CD5^+$ B ≥ 0.05; Supplementary Table 1).

The fraction of CLL patients affected by IPA or TR mutations shown in Fig. 3c, Extended Data Figs. 7c and 8a were calculated as follows: If the CLL-IPA isoform was testable by RNA-seq, all 59 CLL samples were considered. If the CLL-IPA isoform was not being tested by RNA-seq (because, for example, the upstream exon is located too close to the IPA isoform), then only the 13 CLL samples analysed by 3′-seq were taken into account for calculating the fraction of samples with significant expression of the IPA isoform.

Cell lines. B lymphoblastoid cells (BLCL) are Epstein–Barr virus-immortalized human blood B cells[1]. MEC1 cells are malignant B cells from B-prolymphocytic leukaemia and were provided by O.A.-W. Raji and TMD8 cells are malignant B cells from lymphomas and were a gift from H.-G. Wendel. HEK293 and HEK293T cells (embryonic kidney), HeLa cells (cervical cancer) and A549 cells (lung adenocarcinoma) were purchased from ATCC. Wild-type and DICER-knockout HCT116

cells were provided by V. Narry Kim[35]. BLCL, MEC1 and Raji cells were cultured in RPMI with 20% FBS and 1% penicillin–streptomycin. HEK293, HEK293T, HeLa and A549 cells were cultured in DMEM with 10% FBS and 1% penicillin–streptomycin, whereas HCT116 cells were cultured in McCoy's medium with 10% FBS and 1% penicillin–streptomycin.

Western blotting. Cells were lysed on ice for 30 min with RIPA buffer (50 mM Tris pH 7.4, 150 mM NaCl, 1% NP-40, 1% Na-deoxycholate, 1 mM EDTA, 0.05% SDS), containing freshly added proteinase inhibitor cocktail (Thermo Scientific). For MGA, NUP98, SGK223 and DICER immunobloting, cell lysates were run using 3–8% Tris-Acetate NuPAGE gels with Tris-Acetate running buffer (Life Technologies). For CARD11, AKAP10, BAZ1B, SENP1, CUL3 and RIPK1, 4–12% Bis-Tris NuPAGE gels (Life Technologies) were run with MOPS running buffer and all other proteins were run with MES running buffer (Natural Diagnostics). The separated proteins were transferred to nitrocellulose membranes (Bio-Rad, 1620252), blocked with Odyssey Blocking Buffer (Li-Cor, 927-40000) for 1 h at room temperature, followed by incubation with primary antibodies at 4 °C overnight. After two washes using PBS and 0.1% Tween 20 (PBST), the blots were incubated with IRDye-conjugated secondary antibodies for 50 min at room temperature. After one wash with PBST and two washes with PBS, proteins were detected with Odyssey CLx imaging system (Li-Cor).

The following primary antibodies were used: anti-actin (mouse, Sigma, A4700; rabbit, Sigma, A2066), anti-AKAP10 (mouse, clone 51, Santa Cruz Biotechnology, sc-136512), anti-CARD11 (rabbit, Cell Signaling, 4440S), anti-DICER (rabbit, a gift from W. Filipowicz), anti-DNM1L (mouse, Abcam, ab56788), anti-MGA (rabbit, H-286, Santa Cruz Biotechnology, sc-382569), anti-SFRS15 (SCAF4; mouse, Abnova, H00057466-B01), anti-WSTF (BAZ1B; mouse, clone G-5, Santa Cruz Biotechnology, sc-514287), anti-NUP98 (rabbit, Novus Biologicals, NB100-93325), anti-SGK223 (mouse, Santa Cruz Biotechnology, sc-398164), anti-SENP1 (rabbit, Bethyl Labs, A302-927A-T), anti-CUL3 (rabbit, Bethyl Labs, A301-108A-T), anti-PAWR (Abcam ab92590), anti-RIPK1 (Cell Signaling 4926), anti-GAPDH (goat, V-18, Santa Cruz Biotechnology) and anti-WNT5a/b (rabbit, clone C27E8, Cell Signaling 2530). The secondary antibodies used included anti-mouse IRDye 700 (donkey, Rockland Immunochemicals, 610-730-002), anti-rabbit IRDye 680 (donkey, Li-Cor Biosciences, 926-68073), anti-rabbit IRDye 800 (donkey, Li-Cor Biosciences, 926-32213) and anti-mouse IRDye 800 (donkey, Li-Cor Biosciences, 926-32212).

RT–PCR of IPA isoforms. Total RNA was isolated using Tri reagent solution (Invitrogen AM9738) and digested with DNase I (Invitrogen AM1906). RNA was reverse transcribed using the qScript cDNA SuperMix (Quanta Biosciences 101414-106). RT–PCR reactions were carried out using purified Taq polymerase using a 50 °C annealing temperature and 30 s extension at 72 °C. The linear range of amplification was determined by independent PCRs for each primer set. Primers were designed to be intron-spanning and are listed in Supplementary Table 3.

Induction of IPA isoforms. Endogenous U2AF1, U2AF2 and hnRNPC were knocked down using pLKO-puro lentiviral vector-based shRNAs (Sigma). Virus was produced using the helper plasmids pCMV-VSVG and pCMV-dR8.2 and cells were transduced in six-well plates, selected with puromycin (2 μg ml^{-1}) for 5 days and then collected for RT–PCR or western blot analysis.

To induce IPA isoform expression of DICER, an antisense morpholino oligonucleotide (GeneTools) targeting the 5′ splice site of DICER exon 23 was added directly to sub-confluent HeLa cells at the indicated concentrations in the presence of 6 μM EndoPorter-PEG delivery peptide (GeneTools) and harvested at the indicated time points. The control morpholino was used at 12 μM concentration.

Knockdown of CARD11 full-length and IPA isoforms. Isoform-specific shRNA primers were cloned into the TRC2-pLKO-GFP plasmid using KpnI and EcoRI. Lentivirus was produced as described above and centrifuged at 25,000 r.p.m. for 1 h 45 min at 4 °C (Sorvall WX Ultracentrifuge). Pellets were resuspended and dissolved in cold PBS overnight at 4 °C. The virus titre was estimated by transducing wild-type HEK293T cells. The 12-well culture plate was coated overnight with 5 μg ml^{-1} fibronectin. TMD8 cells were spin-infected and cultivated for three days, followed by western blot analysis of FACS-sorted GFP-positive cells.

Constructs. The V5-DICER construct was obtained from J. Mendell. To generate the DICER-IPA expression plasmid, the DICER-IPA cDNA was amplified from BLCL and cloned into the pCK-V5 plasmid using the BamHI and ApaI restriction sites.

The human *MGA* cDNA (Dharmacon, clone BC136659) was used to PCR-amplify the coding region of full-length *MGA* (8,571 nucleotides plus 6 nucleotides of endogenous Kozak sequence) as well as MGA IPA (3,430 nucleotides (end of exon 9) plus GTGAGTATTAA (intronic sequence that will be translated, followed by a stop codon; see Extended Data Fig. 6a)). MGA IPA was cloned into the pcDNA3.1 expression vector (Life Technologies) using NheI and XhoI sites. GFP fused-MGA IPA was generated by inserting MGA IPA downstream of eGFP using the restriction sites BsrGI and XhoI in the pcDNA3.1-GFP vector. MGA was cloned into pcDNA3.1-GFP using Gibson Assembly Cloning (New England Biolabs) from three pieces.

The full-length *FOXN3* mRNA was amplified from BLCL cDNA. To obtain GFP–FOXN3, it was cloned into pcDNA3.1-GFP[36] using BsrGI and XhoI restriction sites. FOXN3 IPA was PCR-amplified from two fragments. Fragment 1 was amplified from BLCL cDNA and corresponds to amino acids 1–180, whereas fragment 2 was amplified from genomic DNA from PBMC and corresponds to the 32 amino acids generated from intronic sequence, followed by a stop codon. FOXN3 IPA was fused with GFP at the C terminus as described above.

Full-length CHST11 was amplified from *BLCL* cDNA, whereas CHST11 IPA was amplified from genomic DNA. Both were fused to GFP at the C terminus as described above. The integrity of all constructs was confirmed by sequencing.

Functional validation of CLL-IPAs. *CARD11 IPA.* To assess NF-κB activation, lentiviral-transduced TMD8 cells (described above) were used. Cells were fixed with 4% formaldehyde at room temperature for 15 min. After two washes with excess PBS, fixed cells were resuspended with ice-cold PBS and permeablized with 90% methanol for 20 min on ice. Cells were then washed with cold PBS twice and resuspended with the incubation buffer (PBS + 0.5% BSA). Cells were aliquoted and incubated with anti-phospho-NF-κB p65 (1:1,500 dilution, Cell Signaling 3033) for 1.5 h at room temperature. Cells were washed with incubation buffer twice and incubated with fluorochrome-conjugated secondary antibody solution (1:10,000 Alexa Fluor 647 A27040, Invitrogen) for 15 min at room temperature. After two washes with incubation buffer, cells were analysed using a FACS Calibur.

DICER IPA. Full-length V5-DICER and V5-DICER IPA were immunoprecipitated from HEK293T cells as described before[16]. In brief, 48 h after transfection, cells were washed with cold PBS and lysed with IP buffer (20 mM Tris-HCl pH 8.0, 150 mM NaCl, 1 mM EDTA, 0.5% NP-40 and 1× EDTA-free protease inhibitor (Thermo Fisher)) for 30 min on ice with occasional vortexing. The cell lysate was then centrifuged at 20,000*g* for 10 min at 4 °C and the supernatant was collected. The cell lysate was incubated with 3 μg of anti-V5 tag antibody (Invitrogen R960-25) for 30 min on ice, then 900 μg of protein G Dynabeads were added and the reaction was rotated for an additional 2 h at 4 °C. After five washes with IP buffer and twice in DICER assay buffer (20 mM Tris-HCl pH 8.0, 100 mM KCl, 0.2 mM EDTA), 90% of beads were resuspended in DICER assay buffer for miRNA cleavage assay and the remaining beads were stored in 2× Laemmli sample buffer (Sigma) for western blot analysis.

The miRNA cleavage assay was performed as described previously[16]. In brief, synthesized pre-miRNA let-7i oligo (Dharmacon) was incubated with immunoprecipitated beads prepared as described above in the enzymatic mixture (10 μl of immunoprecipitated beads in DICER assay buffer, 2 μl of 20 mM $MgCl_2$, 0.2 μl of 0.4 μM pre-miRNA, 0.1 μl of 100 mM DTT, 0.5 μl of RNaseOUT (Invitrogen) and 7.2 μl of RNase-free water) at 37 °C for 30 min with interval mix. The reaction was stopped by chilling samples on ice and analysed by northern blot. To investigate whether DICER IPA acts as a dominant-negative version of full-length DICER, different ratios of V5-DICER and V5-DICER IPA were mixed and tested with respect to miRNA cleavage.

Reaction mixtures (10 μl) were added to 10 μl RNA loading buffer (95% formamide, 0.025% SDS, 0.025% bromophenol blue, 0.025% xylene cyanol FF, 0.5 mM EDTA) and denatured at 95 °C for 5 min followed by quenching on ice. Samples were run on a 15% TBE/Urea gel followed by transfer to a Hybond-N$^+$ nylon membrane (GE Healthcare RPN303B) using a semi-dry transfer apparatus (Hoefer TE70X). After transfer, membranes were briefly dried and then UV cross-linked twice with 1,200 μJ cm^{-2} each cycle. Cross-linked membranes were pre-hybridized for 1 h at 37 °C in ULTRAhyb-Oligo hybridization buffer (Ambion AM8663) in a rotary oven. DNA probes against the intended target RNA were synthesized as oligos and labelled with γ^{32}P-ATP in the presence of T4 polynucleotide kinase (NEB M0201S) for 30 min at 37 °C. Labelled probes were purified through G-25 microspin columns containing Sephadex resin (GE Healthcare 27-5325-01). Membranes were hybridized with labelled probe overnight at 37 °C in a rotary oven. The next day, membranes were washed twice in 2× SSC/0.1% SDS for 5 min each at 37 °C followed by one wash in 0.1× SSC/0.1% SDS for 5 min at 37 °C. Membranes were exposed to phosphorimager screens and scanned.

To assess whether expression of DICER IPA influences miRNA expression in vivo, endogenous let-7 miRNA expression levels were measured by northern blot analysis of total RNA (22 μg) from wild-type and DICER knockout HCT116 cells. DICER knockout HCT116 cells were transfected with different amounts of V5-DICER and V5-DICER IPA. Cells were harvested 3 days after transfection with Lipofectamine 2000 to assess DICER protein expression and corresponding endogenous let-7 levels.

FOXN3 IPA. The fork-head domain of FOXN3 is necessary for transcriptional repression of FOXN3 target genes. Thus, truncation of the fork-head domain predicts derepression of the target genes. Known target genes are *PIM2* and *MYC*[20,37]. MEC1 cells were nucleofected with pcDNA 3.1 vector containing GFP, GFP–FOXN3 or GFP–FOXN3 IPA using SF Cell Line 4D-Nucleofector X Kit (Lonza, Program FF-120). After 48 h, GFP$^+$ cells were FACS sorted, lysed immediately (Cells-to-cDNA II Kit, Ambion) and RNA was extracted. cDNA was synthesized

by qScript cDNA SuperMix (Quanta Biosciences) and quantitative PCR was performed using FastStart universal SYBR green master mix (Roche) on a 7900HT Fast Real-Time PCR System (Applied Biosystems). The experiment was performed from five biologically different replicates.

MGA IPA. Raji cells were nucleofected with pcDNA3.1 vector containing GFP, GFP–MGA or GFP–MGA IPA using Cell Line Nucleofector Kit V (Lonza, Program M-013). After 48 h, GFP$^+$ Raji cells were FACS-sorted and lysed immediately in lysis buffer (Cells-to-cDNA II Kit, Ambion) and RNA was extracted. cDNA synthesis and qRT–PCR was as described for FOXN3. qRT–PCR was done in technical triplicates from three biologically different experiments. MYC target genes were previously published[38,39]. E2F-binding sites in MYC target genes were identified using the Encode Transcription Factor ChIP–seq track, or they were previously described[19,39–41]. T-boxes were described for *ATF4* and *CDKN1B*[42,43].

CHST11 IPA. 3′-seq data were used to identify overexpressed WNT proteins in CLL cells compared to normal B cells. The expression of WNT was validated in MEC1 cells by qRT–PCR. WNT5B was the WNT with the highest expression in MEC1 cells.

For WNT detection in media, MEC1 cells stably expressing GFP, GFP–CHST11 or GFP–CHST11 IPA were counted and washed once with RPMI without FCS. Twenty million cells were cultured in 10 ml RPMI plus 1% pen–strep in one 10-cm culture dish. After 18 h, conditioned medium was collected by centrifugation at 280*g* for 5 min and passed through a 0.45-μm filter. The supernatant was concentrated by an Amicon Ultra-4 centrifugal filter (Millipore, UFC800324) at 3,000*g* at 10 °C for 2 h. The concentrated medium (~50 μl) was collected and subjected to western blot analysis using anti-WNT5a/b antibody (Cell Signaling 2530). The corresponding cell pellets were also collected for western blot analysis.

To assess paracrine WNT activity in MEC1 cells expressing CHST11 IPA, MEC1 cells were nucleofected with pcDNA3.1 vector containing GFP, GFP–CHST11 or GFP–CHST11 IPA. After 24 h, GFP$^+$ cells were FACS sorted and cultivated for three days. The conditioned medium was collected and added to HEK293T cells which were transiently transfected with a WNT reporter plasmid (Addgene 12456, M50, Super 8x TOPFlash) or WNT reporter control plasmid with mutated TCF/LEF binding sites (Addgene 12457, M51, Super 8x TOPFlash mutant)[44]. The conditioned medium was added 24 h after transfection. Luciferase activity was measured 24 h after the addition of conditioned medium using a Glomax 96 Microplate Luminometer as described previously[45].

Intersection of somatic mutations in CLL with IPA. CLL RNA-seq samples ($n = 44$) with available somatic DNA mutation and prognostic data were available to us to map IPA isoform expression[3]. The somatic mutations were obtained using exome sequencing that included extended exon boundaries[46]. We intersected the occurrence of somatic mutations with IPA isoforms in these samples. We focused on truncating mutations (nonsense mutations, frame-shift mutations and splice-site mutations) in expressed genes as they were likely to have a similar outcome as IPA.

The IGVH status of CLL samples was assessed at MSKCC for the CLL samples studied by 3′-seq. The IGVH status of 44 RNA-seq samples was published[3].

Positions of TR mutations. The positions of TR mutations in CLL were obtained from the published CLL somatic mutation datasets[3,7,8]. The positions of TR mutations in solid cancers of TSGs and of genes targeted by CLL-IPAs were obtained from the MSK cbio portal (date of reference, 23 February 2018, containing >86,000 cancer samples with 97% derived from solid tumours)[4]. The position with the highest number of TR mutations was used (hot spot) and is indicated by the symbol. The symbol is lacking if the genes had TR mutations without a hot spot.

Number of amino acids of full-length or IPA-generated truncated proteins. To calculate the number of amino acids of full-length proteins, we used the longest Ref-seq annotated mRNA isoform, obtained the number of coding nucleotides and divided this number by three to obtain the total number of amino acids. To calculate the number of amino acids of the IPA-generated truncated proteins we counted the number of nucleotides from the start codon to the end of the exon located upstream of the IPA isoform and divided this number by three to obtain the number of retained amino acids. This number also provided information about the reading frame of the protein at the exon/intron junction located upstream of the IPA isoform. We then used the correct reading frame and translated the intronic nucleotides until an in-frame stop codon was detected. The amino acids translated from intronic sequence were added to the retained amino acids to obtain the size of the IPA-generated truncated proteins.

The fraction of retained CDR is the number of amino acids retained (up to the end of the exon located upstream of the IPA isoform) divided by the number of amino acids calculated from the longest mRNA isoform encoding the full-length protein.

Identification of known and novel TSGs. For known TSGs, we used the 301 TSGs reported by Davoli et al.[5] that were expressed in CLL samples. Davoli used a computational method (TUSON Explorer) to predict 301 TSGs from genomic sequencing data obtained from more than 8,200 cancers (>90% are derived from solid tumours).

For novel TSGs, we used the data from the MSK cbio portal (see above). It was previously reported that the variable with the highest predictive power for TSGs was the proportion of TR mutations to all mutations[5]. We calculated this ratio for the 190 genes that generated CLL-IPAs in more than 20% of samples and identified a bimodal distribution with a separation point at 12% TR mutations to all mutations. The genes that generated CLL-IPAs in more than 20% of samples and had a TR mutation rate ≥ 12% in the data from MSK cbio portal were called novel TSG candidates (Supplementary Table 2).

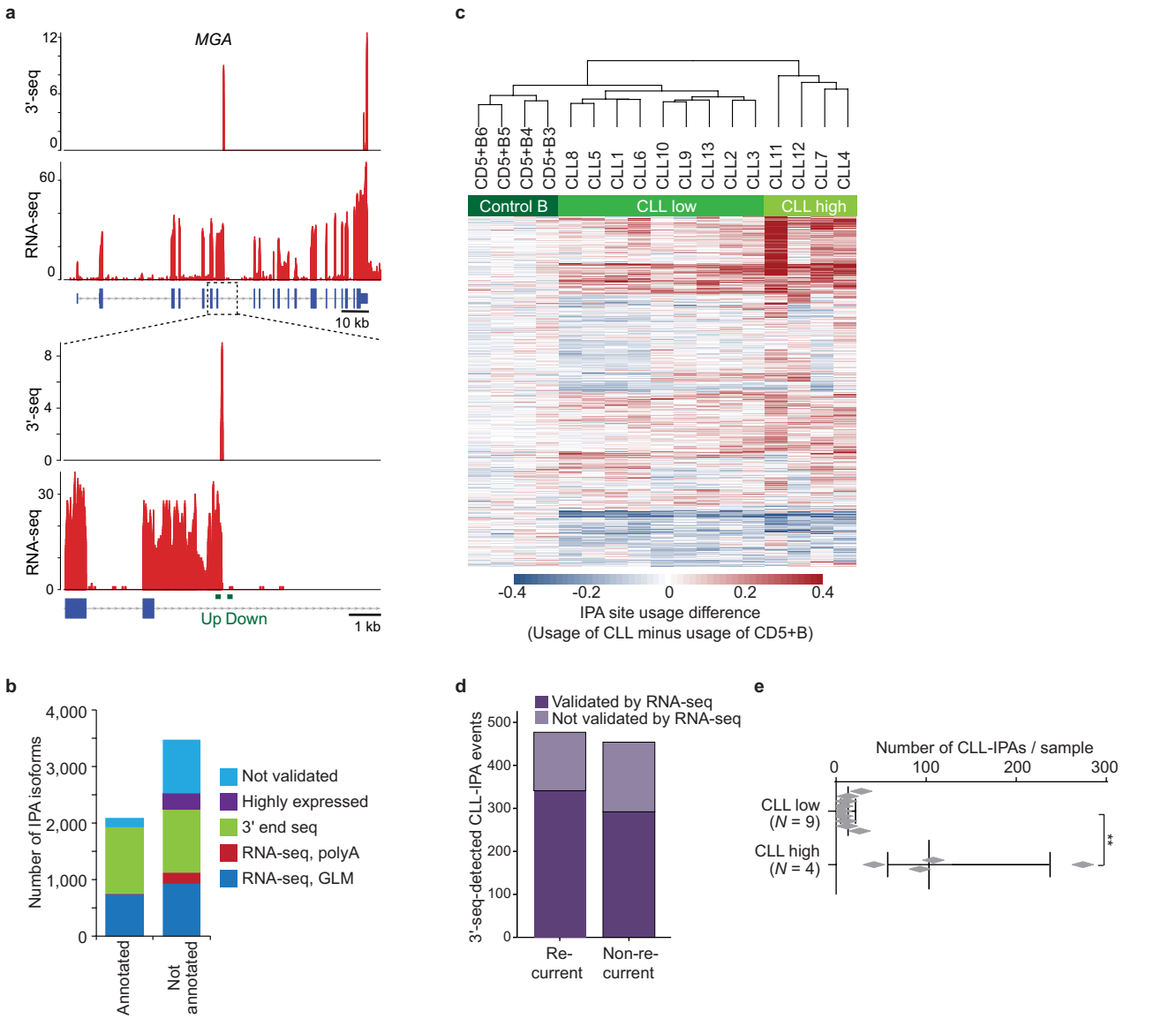
To assess whether known TSGs are enriched among CLL-IPAs a χ^2 test was performed. To exclude that this association occurred by chance, five control lists containing genes with similar coding region length and expression were generated and tested for enrichment of TSGs.

Others statistical methods. To perform enrichment statistics, we used a χ^2 test and calculated the *P* value using a two-sided Fisher's exact test. To assess the functional differences between full-length proteins and IPA-generated truncated proteins (MGA and FOXN3), we used a two-sided *t*-test for independent samples. When comparing three groups (CARD11 and CHST11), a two-sided Kruskal–Wallis test was used. For subsequent pair-wise comparisons, a two-sided Mann–Whitney *U*-test was applied and the *P* values were adjusted with Bonferroni multiple testing correction. For all other tests that assessed the differences of features between two groups, we used a two-sided Mann–Whitney *U*-test. To investigate the spatial relationship between the IPA-generated truncated proteins and hot spot TR mutations, we performed a two-sided Wilcoxon rank-sum test.

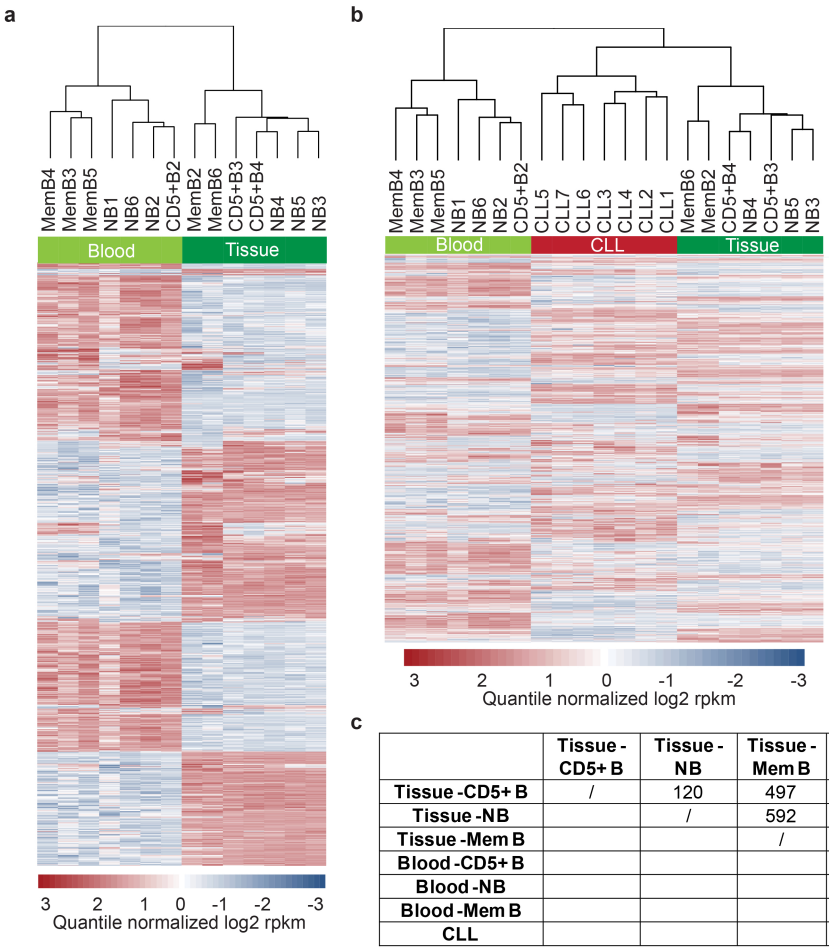
Reporting summary. Further information on research design is available in the Nature Research Reporting Summary linked to this paper.

Data availability. All 3′-seq and RNA-seq data generated and analysed for this study have been deposited in the Gene Expression Omnibus (GEO) database under accession numbers GSE111310 and GSE111793. The code to analyse the data are available at https://bitbucket.org/leslielab/apa_2018/ and the processed data are available in Supplementary Table 1 (for Figs. 1b–d, 2a, 4a, Extended Data Figs. 3 and 4) and Supplementary Table 2 (for Extended Data Fig. 8a), and in the Source Data files (for Figs. 1e, 2c, e, 3a, c, 4b–d, g, Extended Data Figs. 2c, 6j, 7c and 8a). Data on DNA mutations from patients with CLL were provided by D. A. Landau and need to be requested from him. The mutation data on solid cancers were obtained through the MSK cbio portal. The data can be accessed at http://www.cbioportal.org.

29. Béguelin, W. et al. EZH2 is required for germinal center formation and somatic EZH2 mutations promote lymphoid transformation. *Cancer Cell* **23**, 677–692 (2013).
30. Ranzani, V. et al. The long intergenic noncoding RNA landscape of human lymphocytes highlights the regulation of T cell differentiation by linc-MAF-4. *Nat. Immunol.* **16**, 318–325 (2015).
31. Hoek, K. L. et al. A cell-based systems biology assessment of human blood to monitor immune responses after influenza vaccination. *PLoS ONE* **10**, e0118528 (2015).
32. Trimarchi, T. et al. Genome-wide mapping and characterization of Notch-regulated long noncoding RNAs in acute leukemia. *Cell* **158**, 593–606 (2014).
33. Tian, B., Hu, J., Zhang, H. & Lutz, C. S. A large-scale analysis of mRNA polyadenylation of human and mouse genes. *Nucleic Acids Res.* **33**, 201–212 (2005).
34. Anders, S., Reyes, A. & Huber, W. Detecting differential usage of exons from RNA-seq data. *Genome Res.* **22**, 2008–2017 (2012).
35. Kim, Y. K., Kim, B. & Kim, V. N. Re-evaluation of the roles of DROSHA, Export in 5, and DICER in microRNA biogenesis. *Proc. Natl Acad. Sci. USA* **113**, E1881–E1889 (2016).
36. Berkovits, B. D. & Mayr, C. Alternative 3′ UTRs act as scaffolds to regulate membrane protein localization. *Nature* **522**, 363–367 (2015).
37. Karanth, S., Zinkhan, E. K., Hill, J. T., Yost, H. J. & Schlegel, A. FOXN3 regulates hepatic glucose utilization. *Cell Rep.* **15**, 2745–2755 (2016).
38. Li, Z. et al. A global transcriptional regulatory role for c-Myc in Burkitt's lymphoma cells. *Proc. Natl Acad. Sci. USA* **100**, 8164–8169 (2003).
39. Zeller, K. I. et al. Global mapping of c-Myc binding sites and target gene networks in human B cells. *Proc. Natl Acad. Sci. USA* **103**, 17834–17839 (2006).
40. Ren, B. et al. E2F integrates cell cycle progression with DNA repair, replication, and G(2)/M checkpoints. *Genes Dev.* **16**, 245–256 (2002).
41. Taubert, S. et al. E2F-dependent histone acetylation and recruitment of the Tip60 acetyltransferase complex to chromatin in late G1. *Mol. Cell. Biol.* **24**, 4546–4556 (2004).
42. Jenner, R. G. et al. The transcription factors T-bet and GATA-3 control alternative pathways of T-cell differentiation through a shared set of target genes. *Proc. Natl Acad. Sci. USA* **106**, 17876–17881 (2009).
43. Jezkova, J. et al. Brachyury regulates proliferation of cancer cells via a p27Kip1-dependent pathway. *Oncotarget* **5**, 3813–3822 (2014).
44. Veeman, M. T., Slusarski, D. C., Kaykas, A., Louie, S. H. & Moon, R. T. Zebrafish prickle, a modulator of noncanonical Wnt/Fz signaling, regulates gastrulation movements. *Curr. Biol.* **13**, 680–685 (2003).
45. Mayr, C. & Bartel, D. P. Widespread shortening of 3′UTRs by alternative cleavage and polyadenylation activates oncogenes in cancer cells. *Cell* **138**, 673–684 (2009).
46. Landau, D. A. et al. Evolution and impact of subclonal mutations in chronic lymphocytic leukemia. *Cell* **152**, 714–726 (2013).



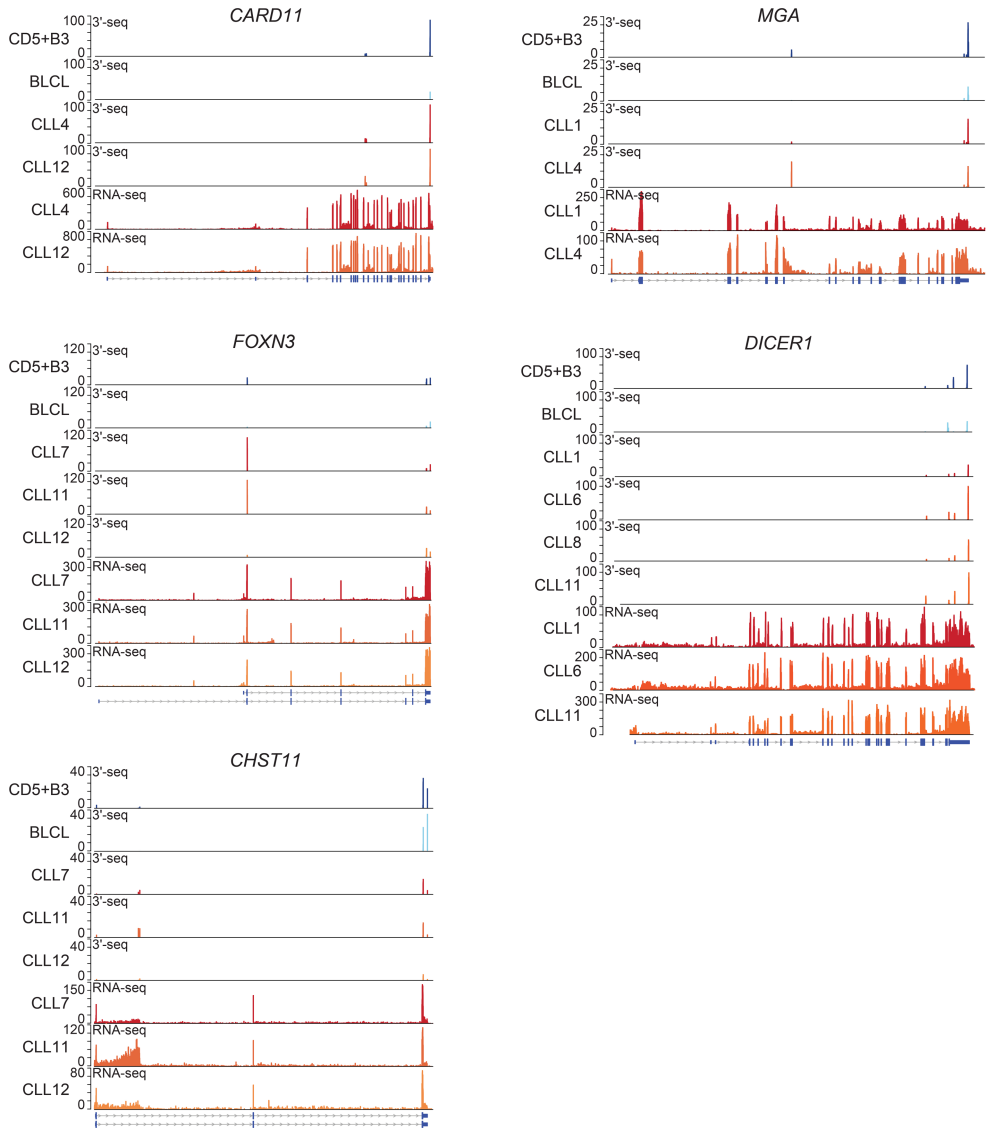
Extended Data Fig. 1 | Validation of IPA isoforms by independent methods and identification of CLL-IPAs used for further analysis. **a**, RNA-seq data were used to validate the presence of IPA isoforms using a GLM. Within two 100-nucleotide windows (green bars) separated by 51 nucleotides and located up- and downstream of the IPA peak, the RNA-seq reads were counted. The IPA peak was considered validated if adjusted $P < 0.1$ (see Methods). Out of $n = 5{,}587$ tested IPA isoforms, $n = 1{,}662$ were validated by this method. Shown is *MGA* as a representative example. **b**, As only a fraction of IPA isoforms were validated by the method from **a**, additional methods were used to obtain independent evidence for the presence of the IPA isoforms. Independent evidence was obtained using untemplated adenosines from RNA-seq data or through the presence of the IPA isoform in other 3′-seq protocols[10]. As the majority of immune cell types used in this study have not been investigated using other 3′-seq protocols and IPA isoform expression is cell type-specific[2], highly expressed IPA isoforms (>10 TPM) were not excluded from further analysis even if no read evidence was found by other protocols. **c**, Hierarchical clustering based on IPA site usage separates the 3′-seq dataset into four groups. It separates $CD5^+$ B from CLL samples and clusters CLL samples into three different groups. Shown is the usage difference of the 20% most variable IPA isoforms across the dataset ($n = 342$). Four out of thirteen CLL samples cluster away from the rest of the samples and are characterized by a high number of IPA isoforms (CLL high). **d**, The GLM (FDR-adjusted $P < 0.1$, IPA usage difference ≥ 0.05, IPA isoform expressed in $CD5^+$ B < 8 TPM) identified 477 recurrent (significantly upregulated in at least 2 out of 13 CLL samples by 3′-seq) and 454 non-recurrent (significantly upregulated in 1 out of 13 CLL samples by 3′-seq). IPAs were validated in an independent RNA-seq dataset containing 46 new CLL samples. Among the recurrent IPAs, 71% of testable IPAs were verified using another GLM (see **a**). Among the non-recurrent IPAs, 64% of testable IPAs were verified. **e**, Plotting the number of CLL-IPAs per sample separates the CLL samples investigated by 3′-seq into two groups: 4 out of 13 samples generate a high number of CLL-IPAs (CLL high, median of CLL-IPAs/sample, $n = 100$, range, 42–274), whereas the rest of the samples generate lower numbers (CLL low, median, $n = 9$, range, 5–28). Centre bar denotes the median; error bars denote the interquartile range. $^{**}P = 0.003$, two-sided Mann–Whitney *U*-test.



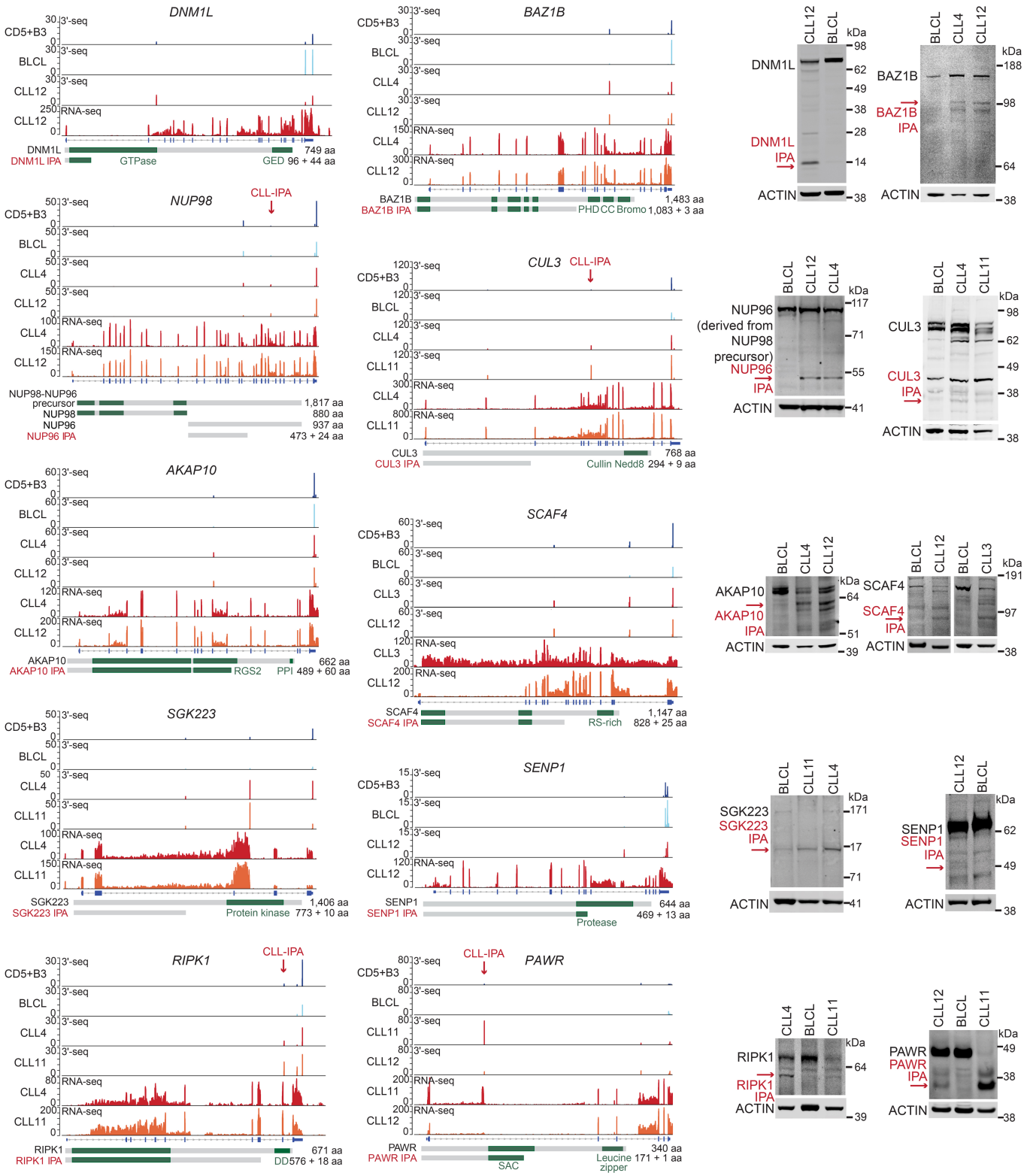
c

	Tissue - CD5+ B	Tissue - NB	Tissue - Mem B	Blood - CD5+ B	Blood - NB	Blood - Mem B	CLL
Tissue -CD5+ B	/	120	497	3,565	1,202	1,381	586
Tissue -NB		/	592	4,805	2,149	2,305	867
Tissue -Mem B			/	2,982	1,620	1,266	679
Blood -CD5+ B				/	3	77	1,605
Blood -NB					/	149	2,211
Blood -Mem B						/	2,298
CLL							/

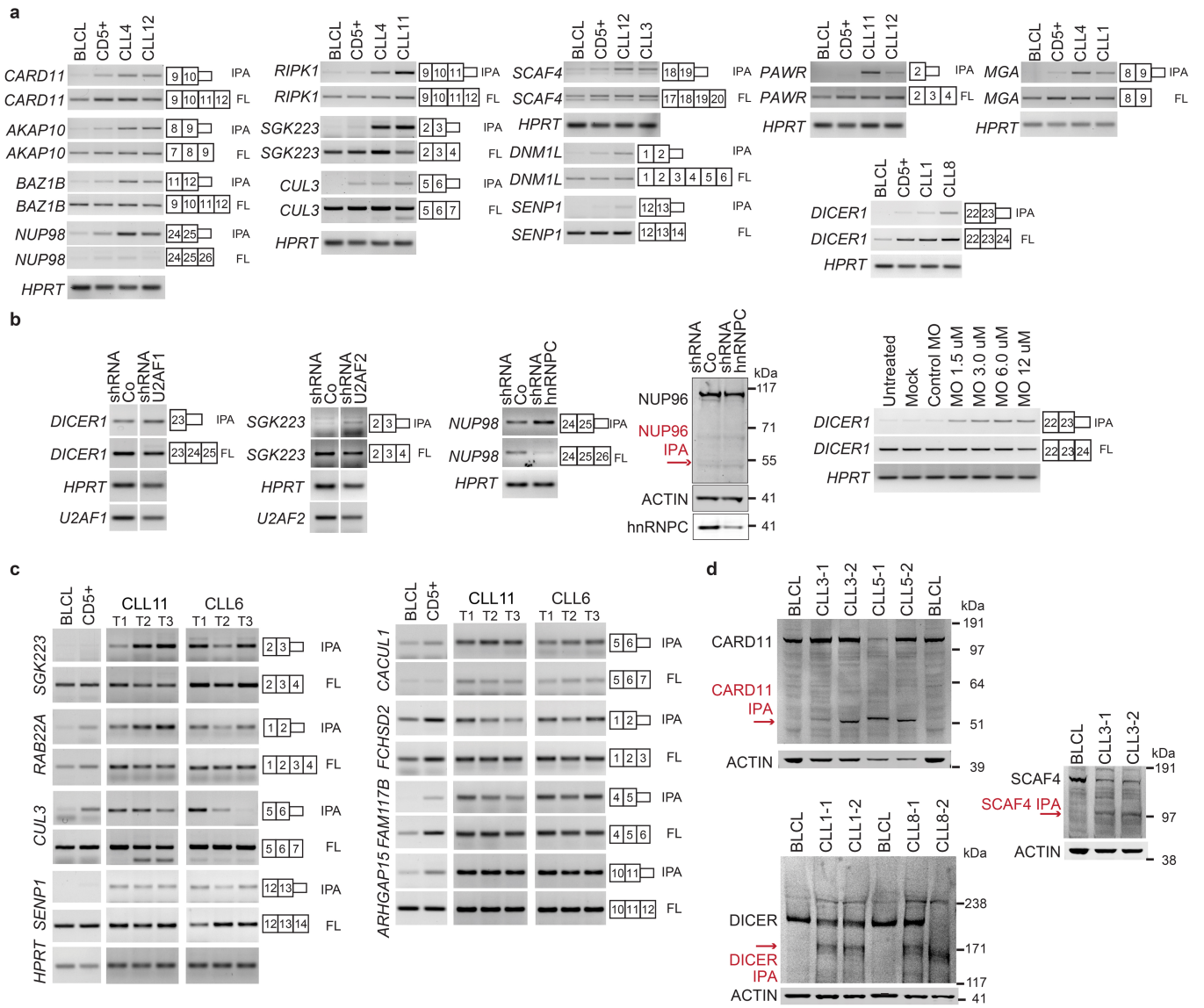
Extended Data Fig. 2 | The normal B cell counterpart of CLL cells are CD5$^+$ B cells derived from lymphoid tissue. a, Hierarchical clustering of normal human B cells (naive B (NB), memory B (MemB) and CD5$^+$ B) derived from lymphoid tissues or peripheral blood based on mRNA expression obtained from RNA-seq. The heat map shows the 20% most variable genes across the dataset ($n = 1{,}887$). The gene expression profiles of B cell subsets derived from peripheral blood or lymphoid tissue differ substantially, although the same markers were used for purification. **b**, As in **a**, but RNA-seq data from CLL samples were added to the analysis. The heat map shows the 20% most variable genes across the dataset ($n = 2{,}078$). CLL samples cluster with tissue-derived and not with blood-derived normal immune cells. **c**, Number of all differentially expressed genes from the analysis shown in **b**.



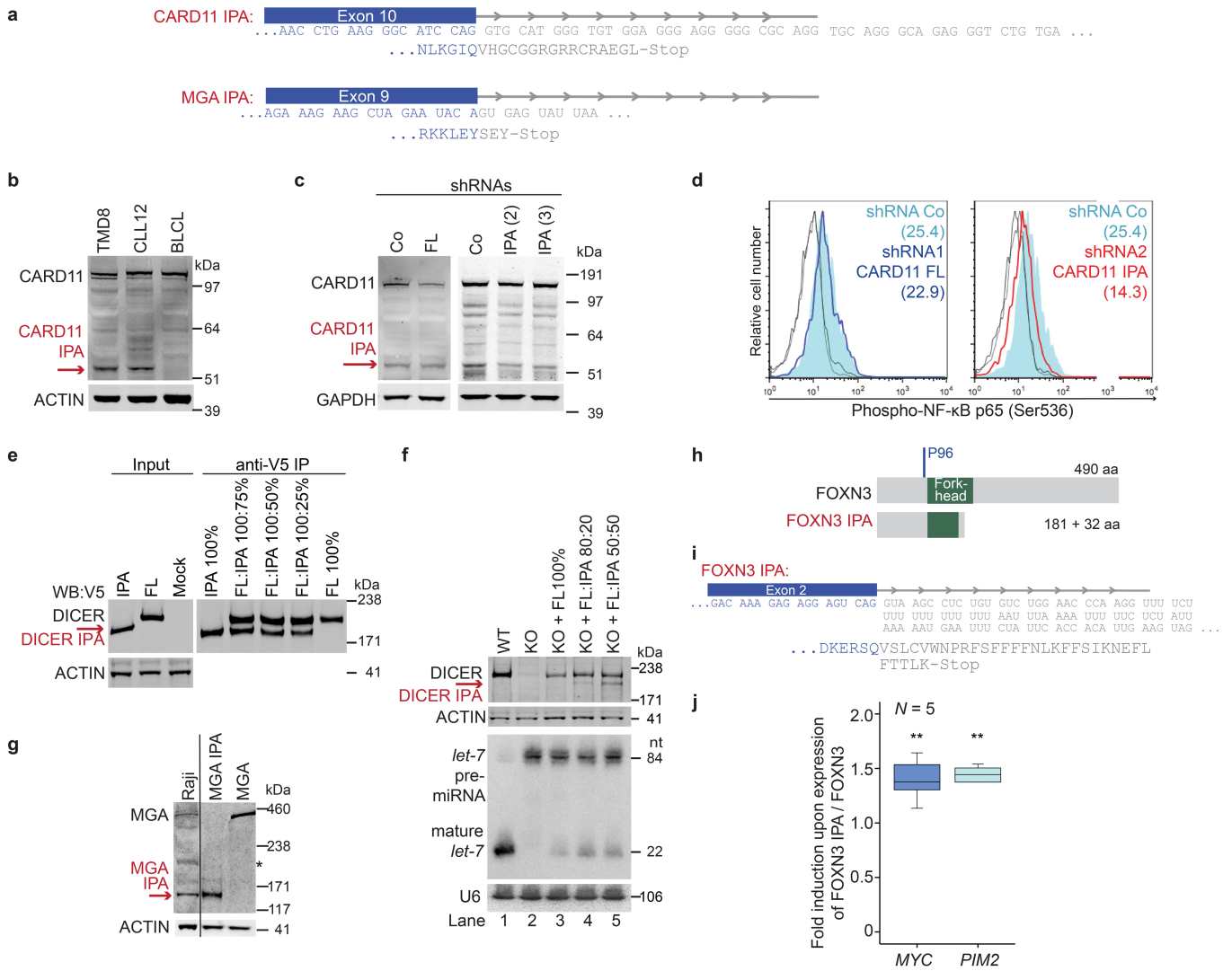
Extended Data Fig. 3 | The 3′-seq and RNA-seq tracks of functionally validated CLL-IPAs. Five CLL-IPAs were functionally validated. Their 3′-seq and RNA-seq tracks are shown here and in Fig. 2a. Data are shown as in Fig. 1b. The corresponding RT–PCRs are shown in Extended Data Fig. 5a.



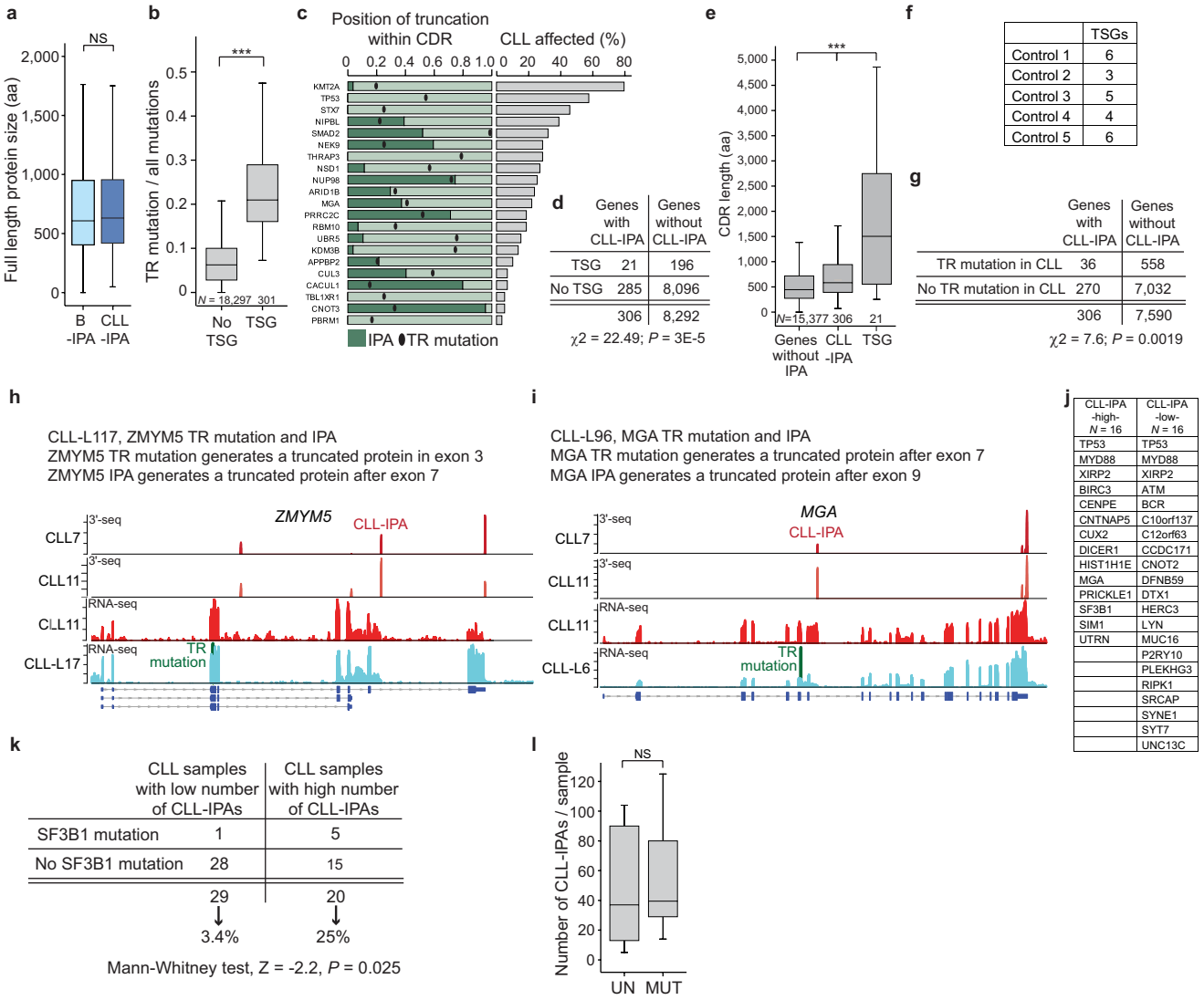
Extended Data Fig. 4 | CLL-IPAs generate truncated mRNAs and proteins. Gene models and western blots of 10 candidates depicted as in Figs. 1b and 2a show that CLL B cells generate full-length and IPA-generated truncated proteins. BLCL were used as control B cells and were included in the 3′-seq tracks. Actin was used as loading control on the same blots. For gel source data see Supplementary Fig. 1.



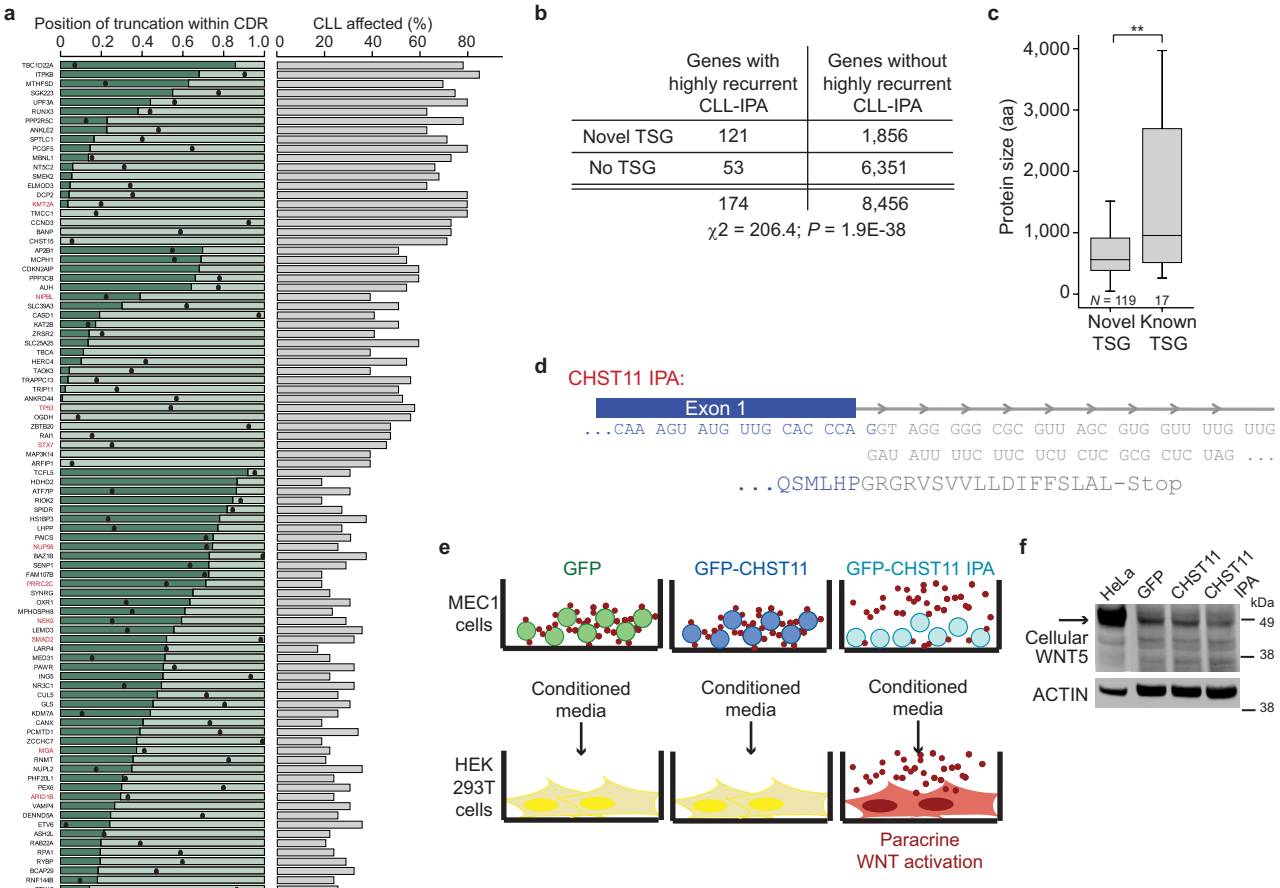
Extended Data Fig. 5 | Validation of the IPA-generated truncated mRNAs and validation of their stable expression over time. a, Detection of full-length and IPA-generated truncated mRNAs by RT–PCR in normal B cells (CD5$^+$ B, BLCL) and CLL cells used in the western blot validations shown in Fig. 2a and Extended Data Fig. 4. All experiments were performed twice with similar results. Primers to amplify the mRNA isoforms are located in the first and last exons shown in the gene models and are listed in Supplementary Table 3. *HPRT* was used as loading control. **b**, Induction of truncated mRNAs and proteins through shRNA-mediated knockdown of splicing factors. All experiments were performed twice with similar results. U2AF1 was knocked down in HeLa cells, U2AF2 was knocked down in HEK293 cells and hnRNPC was knocked down in A549 cells. Shown as in **a**, except for NUP96, which is shown as in Extended Data Fig. 4. NUP96 is derived from NUP98 precursor. Induction of DICER1 IPA by transfection of increasing amounts of anti-sense morpholinos (MO) directed against the 5′ splice site of intron 23 of *DICER1* in HeLa cells. Shown are RT–PCRs. **c**, RT–PCRs, performed once, on expression of full-length and IPA isoforms for eight CLL-IPAs in samples from two patients with CLL and control B cells (CD5$^+$ B, BLCL). The samples were collected over a time interval of over 6 years. CLL11: T1, 17 months after diagnosis, T2, 24 months, T3, 44 months; CLL6: T1, 16 months, T2, 49 months, T3, 91 months (42 months after treatment). Samples from all time points (except CLL6, T3) were obtained from untreated patients. The primers for amplifications of the products were located in the first and last exons shown in the gene models and are listed in Supplementary Table 3. Expression of *HPRT* serves as loading control. The same gel picture of HPRT is shown in Fig. 3b for CLL samples and in **a**, far right panel, for BLCL and CD5$^+$ control samples. All tested CLL-IPA isoforms were detectable at several time points during the course of the disease. Compared with CD5$^+$ B cells, expression of FCHSD2 IPA was not significantly upregulated in CLL. **d**, Western blots of full-length and IPA-generated truncated proteins from CARD11, DICER and SCAF4. All experiments were performed twice with similar results. Actin was used as loading control. Shown are samples from normal B cells (BLCL) and two patients with CLL, both at two different time points 0.5–10 months apart. For gel source data, see Supplementary Fig. 1.



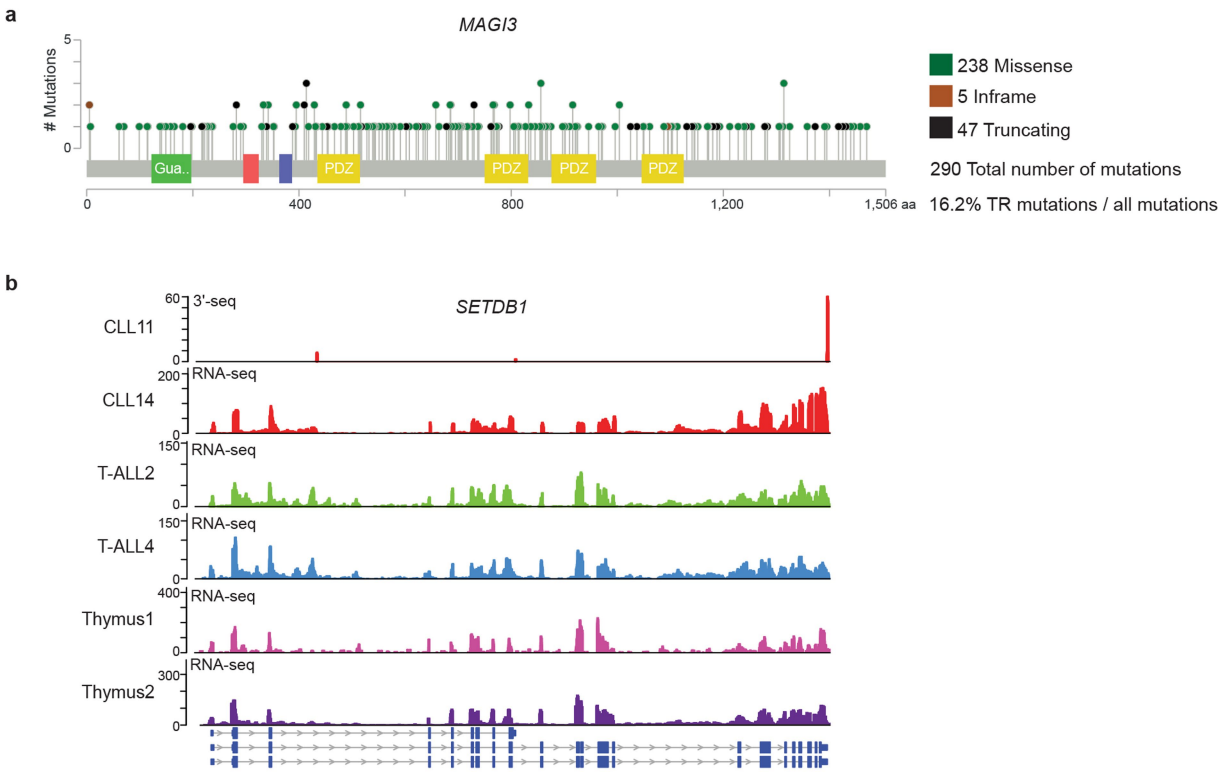
Extended Data Fig. 6 | IPA-generated truncated proteins resemble the protein products of truncating DNA mutations and have cancer-promoting properties. a, CARD11 IPA results in translation of intronic nucleotides (grey) until an in-frame stop codon is encountered. This results in the generation of 16 new amino acids (grey) downstream of exon 10. In the case of MGA IPA, three new amino acids downstream of exon 9 are generated. **b**, Western blot showing that TMD8 cells express similar amounts of CARD11 IPA as CLL samples. The western blot is shown as in Fig. 2a and was performed twice. Actin was used as loading control. **c**, Western blot (as in **b**) showing full-length CARD11 as well as CARD11 IPA in TMD8 cells expressing a control shRNA (Co), an shRNA that exclusively knocks down the full-length protein and two different shRNAs that exclusively knock down the CARD11 IPA isoform. The experiment was performed twice with similar results. GAPDH was used as loading control. **d**, Endogenous phospho-NF-κB p65 levels were measured by FACS in TMD8 cells expressing the indicated shRNAs from **c**. Mean fluorescent intensity values are shown in parentheses in FACS plots of a representative experiment out of three. **e**, Immunoprecipitation of V5-DICER or V5-DICER IPA from HEK293T cells using an anti-V5 antibody. The experiment was performed twice with similar results. 2.5% of input was loaded. **f**, The extent of miRNA processing depends on the expression levels of full-length DICER, but not IPA. Shown are wild-type (WT) and DICER knockout (KO) HCT116 cells. Re-expression of different amounts of full-length DICER1 protein in the knockout cells (measured by western blot of DICER1 in the top panel) results in different levels of endogenous let-7 expression (measured by northern blot in the bottom panel; compare lanes 3 and 4). Expression of DICER IPA has no influence on miRNA processing (compare lanes 4 and 5). Actin and U6 were used as loading controls. The experiment was performed twice with similar results. **g**, Western blot of MGA. MGA and MGA IPA were cloned and expressed in HEK293T cells to confirm the predicted protein size. The experiment was performed twice with similar results. Shown is also the endogenous MGA expression in Raji cells. Actin was used as loading control on the same blot. Asterisk denotes an unspecific band. **h**, Protein models of full-length and FOXN3 IPA are shown as in Fig. 2b. The IPA-generated protein truncates the fork-head domain and is predicted to lose the repressive activity. **i**, As in **a**, but for FOXN3. FOXN3 IPA generates 32 new amino acids downstream of exon 2. **j**, FOXN3 IPA significantly derepresses expression of the oncogenic targets *MYC* and *PIM2*. Fold change in mRNA level of endogenous genes in MEC1 B cells after transfection of GFP–FOXN3 IPA compared with transfection of full-length GFP-FOXN3. *HPRT*-normalized values are shown as box plots (as in Fig. 1e) from $n = 5$ biologically independent experiments, each performed in technical triplicates. $^{**}P = 0.002$, two-sided t-test for independent samples. For gel source data, see Supplementary Fig. 1.



Extended Data Fig. 7 | Inactivation of TSGs by CLL-IPAs independently of DNA mutations. a, The distribution of full-length protein size of genes that generate CLL-IPAs ($n = 306$) and B-IPAs ($n = 2,690$) is shown in amino acids. Box plots are as in Fig. 1e. $P = 0.87$, two-sided Mann–Whitney *U*-test. **b**, TR rate (ratio of TR mutations compared to total mutations) is shown for known TSGs obtained previously[5]. Box plots are as in Fig. 1e. $P = 1 \times 10^{-155}$, two-sided Mann–Whitney *U*-test. **c**, Known TSGs, obtained previously[5] that are targeted by CLL-IPAs ($n = 21$) are shown. Dark green bars indicate the fraction of retained CDRs for each IPA-generated protein. Black dots indicate the hot spot positions of TR mutations obtained from MSK cbio portal. CLL-IPAs mostly occur upstream or within 10% (of overall amino acid length) of the mutations. $P = 0.04$, two-sided Wilcoxon rank-sum test. **d**, Contingency table for enrichment of TSGs among genes that generate CLL-IPAs. *P* value was obtained from two-sided Fisher's exact test. TSGs were obtained previously[5]. **e**, TSGs and genes that generate CLL-IPA isoforms have longer CDRs than genes that do not generate IPA isoforms. Box plots are as in Fig. 1e. $P = 1 \times 10^{-80}$, two-sided Kruskal–Wallis test. **f**, Five control gene lists ($n = 306$, each) with a similar size distribution as CLL-IPAs and expressed in CLL were tested for enrichment of TSGs. Shown is the number of TSGs found. A χ^2 test did not show a significant enrichment of TSGs among the control genes. **g**, Contingency table for enrichment of TR mutation genes in CLL among genes that generate CLL-IPAs. *P* value was obtained from two-sided Fisher's exact test. **h**, ZMYM5 is truncated by a TR mutation and an IPA isoform in the same patient, but the aberrations are predicted to result in different truncated proteins. A 10-bp deletion in exon 3 results in a frameshift leading to the generation of a truncated ZMYM5 protein, whereas ZMYM5 IPA (not yet annotated) produces a truncated protein containing 352 more amino acids in the same patient. The genes shown in **h** and **i** are the only genes with simultaneous presence of a TR mutation and CLL-IPA out of $n = 268$ tested. The position of the TR mutation is indicated in green. CLL7 and CLL11 3′-seq and RNA-seq tracks are shown for comparison reasons. **i**, MGA is truncated by a TR mutation and an IPA isoform in the same patient. The TR mutation affects the 5′ splice site of intron 7, thus generating two additional amino acids downstream of exon 7, whereas the IPA isoform encodes a truncated MGA protein containing three more amino acids downstream of exon 9. Mutation and 3′-seq analysis were performed once. CLL7 and CLL11 are shown for comparison reasons. **j**, Shown are additional recurrent ($n > 1$) DNA mutations found by exome sequencing of CLL patient samples stratified by a high or low number of CLL-IPAs per patient. Only the top and bottom 16 samples with high or low CLL-IPAs are shown to normalize the number of samples analysed. This analysis is only descriptive and no test was performed. **k**, Significant enrichment of SF3B1 mutations in the group of CLL samples with abundant CLL-IPA isoforms. Two-sided Mann–Whitney *U*-test was performed. **l**, Abundance of CLL-IPAs is not associated with IGVH mutational status. Shown is the number of CLL-IPAs per sample for patients with mutated (MUT, $n = 30$) or unmutated (UN, $n = 21$) IGVH genes. Box plots are as in Fig. 1e. $P = 0.4$, two-sided Mann–Whitney *U*-test.



Extended Data Fig. 8 | Novel TSG candidates and validation of CHST11 IPA as cancer-promoting isoform. a, As in Fig. 3c, but shown are known (red gene names) and novel TSG candidates (black gene names) among the abundant CLL-IPAs. CLL-IPAs seem to inactivate these genes as they mostly occur upstream or within 10% (of overall amino acid length) of the mutations. $P = 1 \times 10^{-8}$, two-sided Wilcoxon rank-sum test performed on all 136 TSGs; $P = 1 \times 10^{-8}$, two-sided Wilcoxon rank-sum test performed on the novel TSGs, $n = 119$. Position of the TR mutation was determined using the data obtained from the MSK cbio portal and indicates the hot spot mutation. Right, the fraction of CLL samples affected represents the fraction of CLL samples (out of 59) with significant expression of the IPA isoform. Genes were included if they were affected in at least 20% of samples investigated either by 3′-seq or RNA-seq. **b**, Contingency table for enrichment of novel TSGs among highly recurrent CLL-IPAs. *P* value was obtained from two-sided Fisher's exact test. **c**, TSGs have larger protein sizes. Box plots are as in Fig. 1e. $^{**}P = 0.005$, two-sided Mann–Whitney *U*-test. The increased overall mutation rate of known TSGs correlates with larger protein size. $P = 1 \times 10^{-6}$, Spearman's correlation coefficient, $r = 0.74$. **d**, CHST11 IPA generates 18 new amino acids (grey) downstream of exon 1. **e**, Experimental set-up to measure paracrine WNT activity produced by MEC1 B cells either expressing GFP, GFP–CHST11 or GFP–CHST11 IPA and using a WNT reporter expressed in HEK293T cells. Primary CLL cells and the CLL cell line MEC1 express several WNTs, including WNT5B. In the presence of CHST11 WNT (red dots) binds to sulfated proteins on the surface of WNT producing cells, whereas WNT is secreted into the medium in the presence of CHST11 IPA. WNT-conditioned medium activates a WNT reporter in HEK293T cells. This set-up refers to Fig. 4f, g. **f**, Western blot, performed once, for WNT5 shown as in Fig. 4f, but including HeLa cells as positive control for WNT5 expression. Actin was used as loading control on the same blot.



Extended Data Fig. 9 | Cancer-upregulated IPA isoforms are also detected in breast cancer and T-ALL. a, *MAGI3* is a TSG that is preferentially targeted by IPA in breast cancer[27]. Shown is the mutation profile obtained from MSK cbio portal. **b**, Expression of IPA isoforms in T-ALL detected by RNA-seq. Shown are 3′-seq and RNA-seq tracks of a representative mRNA (out of $n = 101$) from CLL samples, T-ALL samples and normal thymus. The T-ALL RNA-seq data were obtained previously[32]. We detected $n = 381$ IPA isoforms in at least one T-ALL sample, $n = 133$ in at least one thymus sample, $n = 104$ in at least one T-ALL and one thymus sample, and $n = 101$ in at least two T-ALL samples, but not in any of the thymus samples.

Extended Data Table 1 | Samples investigated by 3′-seq and RNA-seq

a

	CLL low vs CLL high	Number of CLL-IPAs	Age at diagnosis	Rai stage at sample collection	WBC count at sample collection	IgVH status	Cyto-genetics	Treated before sample collection	Treated after sample collection	Diagnosis to sample collec-tion (time; mo)	Treatment-free survival (yr)	RNA-seq	3′-seq
CLL1	L	13	62	III	153	UN	Del 11q	N	Y	10	1	Y	Y
CLL2	L	7	62	III	300	NA	Del 17p	N	Y	112	9	Y	Y
CLL3	L	26	54	IV	139	NA	Tri12, t(14;19)	N	Y	84	8	Y	Y
CLL4	H	93	72	0	173	UN	Normal	N	Y	37	4	Y	Y
CLL5	L	11	55	III/IV	193	UN	Tri8, del 13q	N	Y	46	4.5	Y	Y
CLL6	L	28	39	I	137	MUT	Del 13q	N	Y	49	3.3	Y	Y
CLL7	H	108	54	IV	111	MUT	Del 13q	N	Y	108	8	Y	Y
CLL8	L	12	72	III	365	UN	Tri12	N	Y	109	9	N	Y
CLL9	L	5	63	III	200	UN	Del 13q, t(6;19)	N	Y	30	2	N	Y
CLL10	L	11	51	III	77	UN	Del 11q	N	Y	70	6	N	Y
CLL11	H	274	39	0	100	UN	Del 11q, 13q, 14q	N	Y	44	5.5	Y	Y
CLL12	H	42	49	II	178	NA	NA	N	N	240	23.3	Y	Y
CLL13	L	7	66	I	125	UN	Del 11q, del 13q	N	Y	5	0.5	N	Y
CLL14	H	160	45	NA	NA	NA	NA	N	NA	112	NA	Y	N
CLL15	L	49	NA	NA	NA	NA	NA	N	NA	NA	NA	Y	N

N, No; Y, Yes; NA, not analyzed

b

Sample	Derived from	Sample name	Markers for sorting	No. of samples
CD5+B	Tonsil	CD5+B3-CD5+B6	CD5+, CD19+	4
NB	Tonsil	NB3-NB4	CD19+, CD27-	2
NB	Blood	NB1-NB2	CD19+, CD27-	2
MemB	Tonsil	M1-M2	CD19+, CD27+	2
GC	Tonsil	GC1-GC2	CD19+, CD38+	2
PC	BM	PC1-PC3	CD138+	3
T	Blood	T2-T3	CD3+	2

BM, bone marrow

c

Sample	Derived from	Sample name	Markers for sorting	No. of samples
CD5+B	Tonsil	CD5+B3-CD5+B4	CD5+, CD19+	2
CD5+B	Blood	CD5+B2	CD5+, CD19+	1
NB	Tonsil	NB3-NB5	CD19+, CD27-	3
NB	Blood	NB1-NB2, NB6	CD19+, CD27-	3
MemB	Tonsil	M2, M6	CD19+, CD27+	2
MemB	Blood	M3-M5	CD19+, CD27+	3
GC	Tonsil	GC1-GC4	CD19+, CD38+	4
PC	BM	PC4-PC21	CD138+	18

a, CLL sample characteristics. **b**, Normal human immune cells investigated by 3′-seq. **c**, Normal human immune cells investigated by RNA-seq.

LETTER

https://doi.org/10.1038/s41586-018-0453-z

The *NORAD* lncRNA assembles a topoisomerase complex critical for genome stability

Mathias Munschauer[1]*, Celina T. Nguyen[1], Klara Sirokman[1], Christina R. Hartigan[1], Larson Hogstrom[1], Jesse M. Engreitz[1], Jacob C. Ulirsch[1,2,3], Charles P. Fulco[1], Vidya Subramanian[1], Jenny Chen[1,4], Monica Schenone[1], Mitchell Guttman[5], Steven A. Carr[1] & Eric S. Lander[1,6,7]*

The human genome contains thousands of long non-coding RNAs[1], but specific biological functions and biochemical mechanisms have been discovered for only about a dozen[2–7]. A specific long non-coding RNA—non-coding RNA activated by DNA damage (*NORAD*)—has recently been shown to be required for maintaining genomic stability[8], but its molecular mechanism is unknown. Here we combine RNA antisense purification and quantitative mass spectrometry to identify proteins that directly interact with *NORAD* in living cells. We show that *NORAD* interacts with proteins involved in DNA replication and repair in steady-state cells and localizes to the nucleus upon stimulation with replication stress or DNA damage. In particular, *NORAD* interacts with RBMX, a component of the DNA-damage response, and contains the strongest RBMX-binding site in the transcriptome. We demonstrate that *NORAD* controls the ability of RBMX to assemble a ribonucleoprotein complex—which we term *NORAD*-activated ribonucleoprotein complex 1 (NARC1)—that contains the known suppressors of genomic instability topoisomerase I (TOP1), ALYREF and the PRPF19–CDC5L complex. Cells depleted for *NORAD* or RBMX display an increased frequency of chromosome segregation defects, reduced replication-fork velocity and altered cell-cycle progression—which represent phenotypes that are mechanistically linked to TOP1 and PRPF19–CDC5L function. Expression of *NORAD* in *trans* can rescue defects caused by *NORAD* depletion, but rescue is significantly impaired when the RBMX-binding site in *NORAD* is deleted. Our results demonstrate that the interaction between *NORAD* and RBMX is important for *NORAD* function, and that *NORAD* is required for the assembly of the previously unknown topoisomerase complex NARC1, which contributes to maintaining genomic stability. In addition, we uncover a previously unknown function for long non-coding RNAs in modulating the ability of an RNA-binding protein to assemble a higher-order ribonucleoprotein complex.

NORAD stands out among long non-coding RNAs (lncRNAs) because it (1) is highly conserved relative to other lncRNAs, (2) is abundantly expressed in many cell types, (3) is upregulated upon DNA damage and (4) induces chromosomal instability and aneuploidy when deleted. This phenotype is intriguing as little is known about the roles of lncRNAs in maintaining a stable genome. A model for lncRNA function suggests that lncRNAs can serve as assembly scaffolds for ribonucleoprotein complexes[6,7], yet this model has been explored in only a few cases. The mechanism that connects the *NORAD* lncRNA to chromosomal instability remains unknown.

Two recent studies have reported PUMILIO, a highly abundant cytoplasmic RNA-binding protein with no known role in genomic stability, as the sole *NORAD*-interacting protein[8,9]. However, these results were obtained from in vitro mixing of exogenous *NORAD* fragments with cytoplasmic extracts, which may not accurately represent the protein contacts of *NORAD* in living cells (Supplementary Note 1).

To reveal the direct interactions of *NORAD* with proteins in live cells, we captured and identified *NORAD*-interacting proteins by combining RNA antisense purification (RAP) with quantitative liquid chromatography–mass spectrometry using isobaric mass tag quantification (RAP MS) (Fig. 1a). HCT116 colon carcinoma cells were treated with 365-nm light after 4-thiouridine labelling[10], which covalently crosslinks proteins to RNA but not to other proteins. lncRNA–protein complexes were purified by RNA hybrid selection with antisense oligonucleotides that target *NORAD*, under denaturing and reducing conditions at high temperature to minimize the co-purification of indirectly bound proteins[2] (Fig. 1a). To identify specific interactors with *NORAD*, we quantitatively compared the resulting proteins to those captured in purifications with antisense oligonucleotides that target the well-characterized 'RNA component of mitochondrial RNA processing endoribonuclease' (*RMRP*), which is not expected to interact with the same proteins as *NORAD* [11]. We analysed biological replicate purifications in a single 4-plex iTRAQ cassette, quantifying 1,361 proteins that each had more than two unique peptides (Fig. 1b). The control purification captured about 85% of *RMRP* transcripts (Extended Data Fig. 1) and enriched the target RNA approximately 550-fold versus input RNA. We found 12 strongly enriched proteins (mean $\log_2$(iTRAQ ratio (NORAD/RMRP)) < -1.6, $P < 0.05$, moderated *t*-test) (Fig. 1c), including 8 of the 10 known core components of the *RMRP* complex[11] and one previously identified candidate rRNA- and/or tRNA-processing factor[12].

We then analysed *NORAD* antisense purifications. Experiments captured 82% of endogenous *NORAD* (Extended Data Fig. 1) (about 80-fold enrichment versus input RNA). We reproducibly identified 45 proteins that met our enrichment criteria (mean $\log_2$(iTRAQ ratio (NORAD/RMRP)) > 1.6, $P < 0.05$, moderated *t*-test) (Fig. 1b). This set of proteins is highly specific to *NORAD*, in that 41 out of the 45 proteins (Fig. 1c) were not among 219 promiscuous binders (Supplementary Note 3). The RNA-binding protein PUMILIO2 (PUM2) was indeed present in our *NORAD* interactome, but it ranked 185th out of the 265 proteins we detected (mean $\log_2$(iTRAQ ratio (NORAD/RMRP)) > 0.5) and did not meet our cut-off for strongly enriched proteins.

Notably, many of the 41 *NORAD*-interacting proteins have key roles in nuclear processes such as DNA unwinding, replication and repair (including PURA, PURB, TAF15, ALYREF, SFPQ, SRSF1, RBM14, DDX17, RBMX and its retrogene RBMXL1). Twenty-nine (71%) of the forty-one proteins localize to the nucleus, nucleoplasm or chromatin, whereas only two (5%) localize exclusively to the cytoplasm (Fig. 1d). The interactome thus points towards an important nuclear function of *NORAD*.

[1]Broad Institute of MIT and Harvard, Cambridge, MA, USA. [2]Division of Hematology/Oncology, Boston Children's Hospital and Department of Pediatric Oncology, Dana-Farber Cancer Institute, Harvard Medical School, Boston, MA, USA. [3]Program in Biological and Biomedical Sciences, Harvard University, Cambridge, MA, USA. [4]Division of Health Sciences and Technology, MIT, Cambridge, MA, USA. [5]Division of Biology and Biological Engineering, California Institute of Technology, Pasadena, CA, USA. [6]Department of Biology, MIT, Cambridge, MA, USA. [7]Department of Systems Biology, Harvard Medical School, Boston, MA, USA. *e-mail: mathias@broadinstitute.org; eric@broadinstitute.org

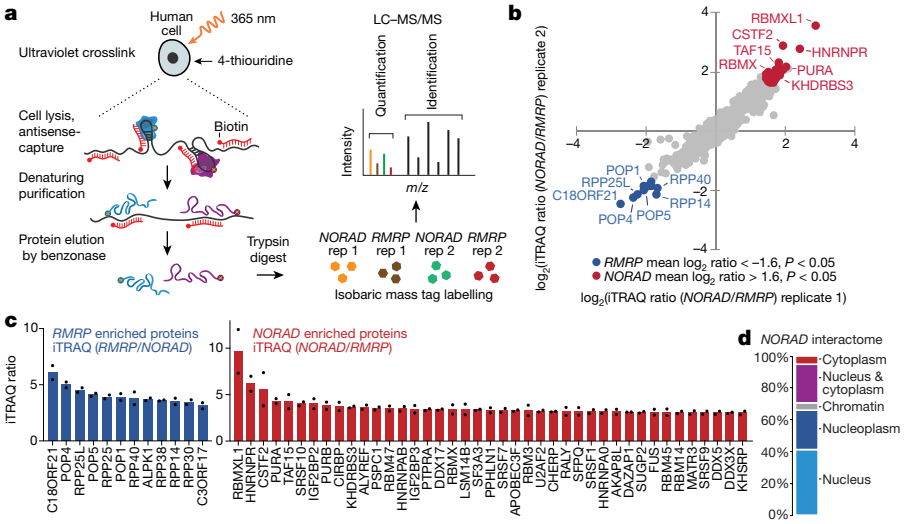


Fig. 1 | *NORAD* directly binds many nuclear proteins in living cells. **a**, Schematic overview of RAP MS. **b**, Quantification of *NORAD*- and *RMRP*-interacting proteins. Scatter plot of $\log_2$-transformed iTRAQ ratios from two biological replicates is shown. Adjusted *P* value, two-tailed moderated *t*-test (Supplementary Note 2). Rep, replicate. **c**, iTRAQ ratios of *RMRP*- (blue) and *NORAD*- (red) enriched proteins. Columns represent the mean of two biological replicate experiments, individual data points are shown (Supplementary Tables 1, 2). **d**, Subcellular localization of *NORAD*-interacting proteins.

Given the overrepresentation of nuclear proteins, we used single-molecule RNA fluorescent in situ hybridization (smRNA FISH) to assess the subcellular localization of *NORAD* in intact cells. In contrast to previous reports that characterized *NORAD* as being located exclusively in the cytoplasm[8], we found that on average 40–50% of *NORAD* transcripts in HCT116 cells reside in the nucleus (Fig. 2a and Extended Data Fig. 2a, b). We confirmed the nuclear localization by subcellular fractionation and quantitative PCR with reverse transcription (RT–qPCR) (Extended Data Fig. 2c, d). Notably, when cells were challenged with DNA damage and replication stress, *NORAD* was upregulated (Extended Data Fig. 2e) and its nuclear localization increased markedly (to about 85%), whereas the localization patterns of control RNAs were unaffected (Fig. 2a, b). Given this shift in localization, we performed RAP experiments with and without DNA damage to confirm that the interactions of *NORAD* with several candidate binders also occur under conditions of DNA damage (Extended Data Fig. 2f, g).

Among the *NORAD*-interacting proteins, we focused on RBMX, the knockdown phenotype of which (impaired DNA damage repair[13] and premature sister-chromatid separation[14]) is closely related to the previously reported *NORAD* knockout phenotype[8]. To explore this connection, we quantified the frequency of chromosome-segregation defects in response to depletion of *NORAD* or RBMX by imaging mitotic cells. We achieved >90% reduction in *NORAD* expression (estimated by RT–qPCR, RNA-sequencing and smRNA FISH) by CRISPR interference (KRAB–dCas9) targeted to the *NORAD* promoter (Fig. 2a and Extended Data Fig. 3a, b). For both wild-type and knockdown cells, we imaged 100 anaphase nuclei and calculated the frequency of DAPI-positive anaphase bridges. Consistent with previous reports[8], *NORAD* depletion caused a significant increase (2.2-fold) in segregation defects (Fig. 2c, d). Importantly, these defects were rescued by expression of full-length *NORAD* in *trans* (Fig. 2d), indicating that the defects are dependent on the *NORAD* RNA. Depletion of RBMX (Extended Data Fig. 3a) caused a comparable increase (2.6-fold) in the frequency of anaphase bridges (Fig. 2d). By contrast, depletion of the cytoplasmic protein PUM2 (Extended Data Fig. 3a) caused no substantial increase in segregation defects (Extended Data Fig. 3c). We reasoned that the interaction between *NORAD* and RBMX may hold important mechanistic insights into *NORAD* function.

To explore this interaction, we mapped RBMX-binding sites on *NORAD* by crosslinking and immunoprecipitation (CLIP). We covalently coupled proteins to RNA using ultraviolet crosslinking[10] and immunopurified RBMX with a specific antibody. We isolated and sequenced RNA crosslinked to RBMX. RBMX displays unusually strong and specific binding to the 5′ end of *NORAD* (Fig. 2e). The RBMX-binding site in *NORAD* extends over more than 800 nucleotides and covers about 15% of *NORAD*—making it eight times larger than the majority of RBMX-binding sites (Extended Data Fig. 3d) and the strongest RBMX-binding region in the transcriptome (Fig. 2f). This unusual binding pattern suggests that *NORAD* serves as a high-affinity binding target for RBMX and contains many RBMX-binding sites. A multiple sequence alignment of *NORAD* transcripts, which was assembled de novo from RNA-sequencing data from 11 mammalian species (Extended Data Fig. 3e), suggests that the RBMX-binding region in *NORAD* is transcribed and conserved throughout mammalian evolution. Next, we performed CLIP for three additional RNA-binding proteins and showed that the RBMX-binding region does not bind PUMILIO, FUBP1 or FUBP3 (Extended Data Fig. 3f).

To confirm that the *NORAD*–RBMX interaction occurs in the nucleus, we performed RBMX RNA immunoprecipitation (RIP) in nuclear and cytoplasmic extracts and showed that over 99% of the total RBMX RIP signal is indeed nuclear (Extended Data Fig. 3g). Consistent with this result, immunofluorescence microscopy suggests that RBMX localizes exclusively to the nucleus (Extended Data Fig. 3h). Finally, depletion of RBMX did not affect subcellular localization of *NORAD* (Extended Data Fig. 3i).

We speculated that *NORAD* might use its large RBMX-binding site to assemble a ribonucleoprotein complex. To examine the role of *NORAD* in such a complex, we sought to identify proteins that bind RBMX and determine whether their interaction with RBMX was dependent on *NORAD*. We performed co-immunoprecipitation and mass spectrometry (co-IP MS) experiments and compared the quantitative enrichment of RBMX-interacting proteins in cells with and without *NORAD* knockdown (Fig. 3a). Importantly, we used a nonspecific RNA and DNA nuclease (benzonase) to ensure that RBMX-binding is direct, rather than being mediated by RNA.

Among the top 11 proteins that bound to RBMX only in the presence of *NORAD*, 7 are linked to DNA replication or repair (Fig. 3b).

Six of these proteins (TOP1, TOP1MT, PRPF19, CDC5L, BCAS2 and MEPCE) were not detected in *NORAD* RAP MS data or were not among the top 200 enriched proteins, which suggests that they bind directly to RBMX and do not interact strongly with *NORAD* (Extended Data Fig. 4a). We further confirmed by western blot the absence of TOP1 in *NORAD* antisense purifications (Extended Data Fig. 4b) and showed that levels of TOP1, RBMX, PRPF19 and CDC5L proteins

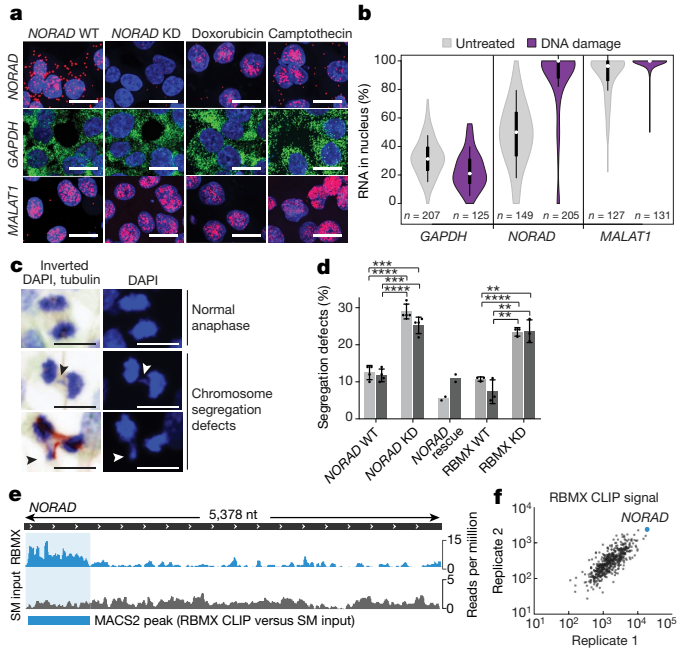


Fig. 2 | *NORAD* and RBMX are present in the nucleus, affect chromosome segregation and interact via a distinct sequence element. **a**, smRNA FISH of *NORAD*, *GAPDH* and *MALAT1* in wild-type (WT), knockdown (KD) and DNA-damaged cells. *GAPDH*, cytoplasmic reference; *MALAT1*, nuclear reference. Scale bar, 20 μm. Images are representative of one experiment; three independent experiments were performed. **b**, Quantification of smRNA FISH experiments. Circles, medians; box limits, 25th and 75th percentiles; whiskers, 1.5× interquartile range; polygons, extreme values. **c**, Normal anaphase nuclei and chromosome segregation defects in *NORAD* knockdown cells. Scale bar, 5 μm. Images are representative of four independent experiments. **d**, Quantification of chromosome segregation defects. One hundred anaphase nuclei are scored for each condition. Grey bars represent different single-guide RNAs. Values are mean ± standard deviation (*NORAD*, $n = 4$, *NORAD* rescue, $n = 2$, RBMX, $n = 3$). $^{****}P < 0.0001$, $^{***}P < 0.001$, $^{**}P < 0.01$, two-tailed Student's *t*-test. **e**, RBMX CLIP data plotted across *NORAD* RNA. SM input, size-matched input for proteins migrating at the same molecular weight range as RBMX. MACS2-identified RBMX peak is shown. **f**, CLIP signal for all RBMX peaks enriched over SM input in two biological replicates.

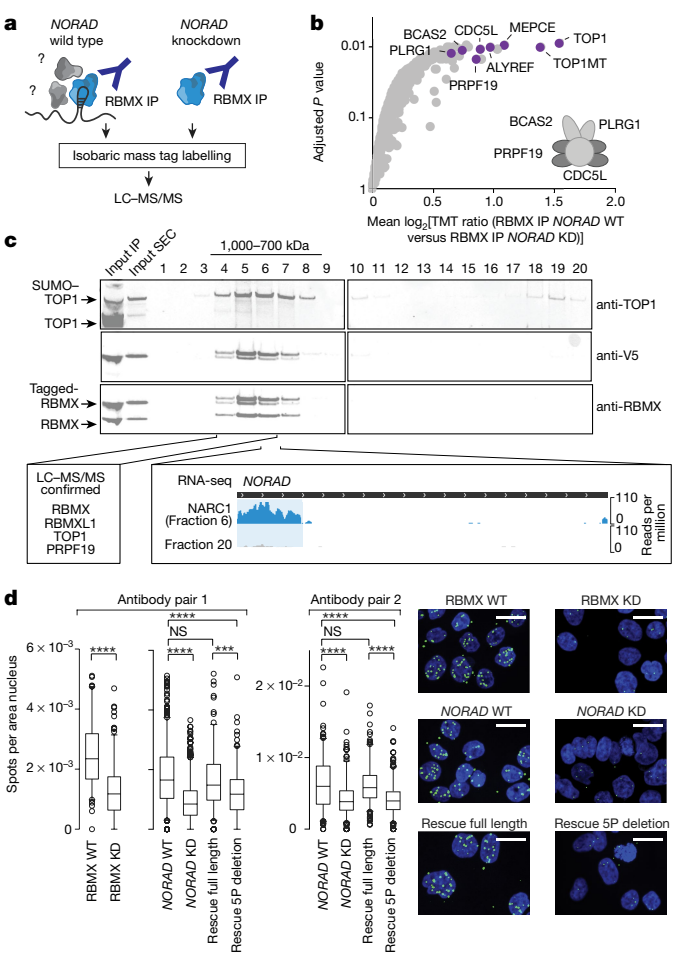


Fig. 3 | *NORAD* modulates RBMX and is essential for assembly of a topoisomerase complex. a, Illustration of RBMX co-IP MS experiments. **b**, Quantification of RBMX-interacting proteins in wild-type versus *NORAD* knockdown cells. Mean $\log_2$ TMT ratios from two biological replicates are shown. Adjusted *P* value, two-tailed moderated *t*-test (Supplementary Note 2). Purple, *NORAD*-dependent RBMX interactors functionally linked to DNA replication and repair. Inset, illustration of PRPF19–CDC5L complex. **c**, Western blot of Flag–RBMX–V5 co-IP followed by size-exclusion chromatography. Fractions 4–6 were pooled for mass spectrometry (Supplementary Table 5). Fraction 6 and fraction 20 were used for RNA sequencing. Western blots are representative of one experiment; three independent experiments were performed. SEC, size-exclusion chromatography. **d**, Proximity ligation assay for RBMX and TOP1. Two different antibody pairs were used. Centre lines, medians; box limits, 25th and 75th percentiles; whiskers, 5th and 95th percentiles; dots, outliers. $^{****}P < 0.0001$, $^{***}P < 0.001$, NS, not significant, two-tailed Mann–Whitney *U* test. Sample size antibody pair 1: RBMX WT, $n = 102$; RBMX KD, $n = 193$; *NORAD* WT, $n = 521$; *NORAD* KD, $n = 559$; rescue full length, $n = 176$; rescue 5P(5′ RBMX-binding site) deletion, $n = 171$. Sample size antibody pair 2: *NORAD* WT, $n = 209$; *NORAD* KD, $n = 232$; rescue full length $n = 211$; rescue 5P deletion, $n = 290$. Representative images are shown. Scale bar, 20 μm.

were not changed upon *NORAD* depletion (Extended Data Fig. 4c). PRPF19, CDC5L and BCAS2, together with PLRG1, make up the core of the human PRPF19–CDC5L complex, and both PRPF19–CDC5L and TOP1 have important roles in DNA replication and genomic stability, as previously reviewed[15,16]. TOP1 suppresses genome instability by preventing interference between replication and transcription[17]. This involves relieving torsional stress in DNA (that is, supercoiling) and suppressing the accumulation of RNA–DNA hybrids (R-loops)[18]. Both R-loops and supercoiled DNA impair replication-fork progression and can lead to genomic instability[17,19]. Stalled replication forks activate the DNA-damage response though ATR signalling. CDC5L binds and activates ATR[20], and the E3 ligase PRPF19 ubiquitylates RPA, enhancing ATRIP–ATR recruitment to stalled replications forks[21].

The precise roles of the remaining two *NORAD*-dependent RBMX interactors in maintaining genomic stability are less well understood. MEPCE binds to the 5′ cap of 7SK[22] and was reported in several studies that aimed to identify proteins involved in the DNA damage response[13,23]; however, its exact function in this process remains unknown. Unlike the six proteins above that were not found by RAP to interact strongly with *NORAD*, a seventh protein—ALYREF—was identified as a strong *NORAD* binder. ALYREF is part of the human TREX complex and interacts with the 5′ end of many RNAs, including *NORAD* (Extended Data Fig. 4d), to facilitate RNA export from the nucleus[24]. ALYREF contributes to genomic stability by suppressing R-loops[24], as does TOP1.

We performed reciprocal co-IP and western blots to confirm that TOP1, ALYREF and CDC5L interact with RBMX and also contact each other (Extended Data Fig. 4e), suggesting that these proteins may constitute a complex. To test whether such a complex exists, we generated cell lines that express epitope-tagged RBMX and performed co-IP experiments (using benzonase to digest unprotected RNA and DNA) followed by native elution and size-exclusion chromatography. Western blot analysis of size-fractionated co-IP samples showed that RBMX,

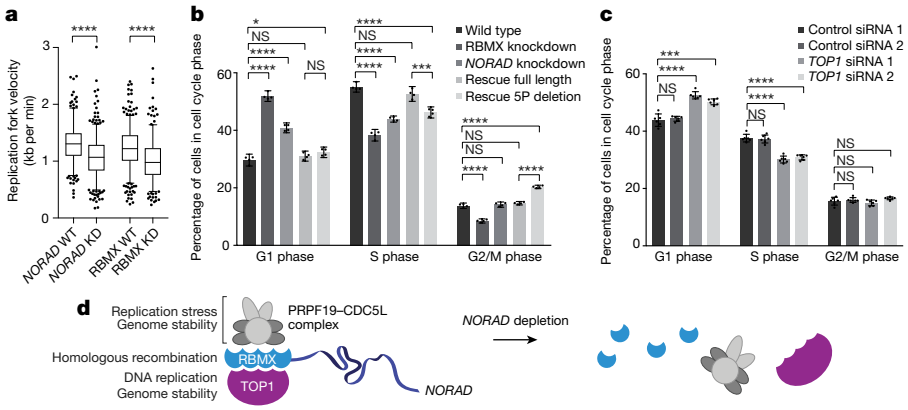


Fig. 4 | Depletion of *NORAD* and NARC1 components affects replication-fork velocity and cell-cycle progression. a, Replication-fork velocity measured by DNA combing (Supplementary Tables 8–10). Centre lines, medians; box limits, 25th and 75th percentiles; whiskers, 5th and 95th percentiles; dots, outliers. $^{****}P < 0.0001$, two-tailed Mann–Whitney *U* test. Sample size: *NORAD* WT, $n = 280$; *NORAD* KD, $n = 428$; RBMX WT, $n = 438$; RBMX KD, $n = 296$. **b**, Cell-cycle analysis by flow cytometry. CRISPR interference was used to deplete RBMX and *NORAD* ($n = 4$; Supplementary Table 11). Fluorescence-activated cell sorting histograms are shown in Extended Data Fig. 5e. **c**, As in **b**, but for TOP1. RNA interference was used to deplete TOP1 ($n = 6$; Supplementary Table 12). Fluorescence-activated cell sorting histograms are shown in Extended Data Fig. 5f. Values are mean ± 95% confidence interval. $^{****}P < 0.0001$, $^{***}P < 0.001$, $^{*}P < 0.05$, NS, not significant, two-tailed Welch's *t*-test. **d**, Model illustrating *NORAD* function.

TOP1 and PRPF19 are part of a 700–1,000-kDa complex (Fig. 3c and Extended Data Fig. 4f). The majority of TOP1 in this complex displays an approximately 50-kDa size shift, consistent with a known SUMO-1 modification of TOP1 proteins that are associated with transcriptionally active or replicating chromatin[25]. Mass spectrometry confirmed that—in addition to RBMX, TOP1 and PRPF19—RBMXL1, which is encoded by an *RBMX* retrogene, is a component of this complex (Fig. 3c). Finally, we speculated that this complex protects *NORAD* from benzonase digestion. We constructed sequencing libraries using RNA extracted from various size-exclusion chromatography fractions. Notably, RNA footprints that matched the previously identified RBMX-binding site in *NORAD* were present only in fractions that contained the complex (Fig. 3c). These data demonstrate that *NORAD* is a physical part of the captured complex.

We next used proximity ligation assays to show that the RBMX–TOP1 interaction occurs in the nucleus, is disrupted by *NORAD* depletion and is rescued by re-introducing full-length *NORAD* into *NORAD*-depleted cells (Fig. 3d and Extended Data Fig. 4g). Importantly, rescue is strongly impaired if the rescue construct lacks the RBMX-binding region (Fig. 3d and Extended Data Fig. 4g).

Our results indicate that *NORAD* modulates the ability of RBMX to interact with other proteins that appear not to bind *NORAD* directly—namely, TOP1 and the core PRPF19–CDC5L complex. Given the distinct molecular composition of this *NORAD*-dependent RBMX complex and the functional importance of its components, we name it *NORAD*-activated ribonucleoprotein complex 1 (NARC1).

Many NARC1 components have prominent roles in maintaining genomic stability. Although individual components such as RBMX or PRPF19 have been reported to contribute to mRNA splicing[13,16], we did not observe global changes in mRNA splicing upon *NORAD* depletion in HCT116 cells (Extended Data Fig. 5a). In other cell types, RBMX and CDC5L can influence the expression of BRCA2 and BRCA1[13,26]. However, *BRCA1* and *BRCA2* were not among differentially expressed genes (Supplementary Table 7) and their proteins levels were not noticeably different in *NORAD*-depleted and normal HCT116 cells (Extended Data Fig. 4c).

Given the essential role of *NORAD* in assembling NARC1, we speculated that *NORAD* depletion may cause a TOP1-related phenotype and directly or indirectly alter DNA replication, which can lead to chromosome segregation defects and genomic instability[27,28]. To assess the functional consequence of *NORAD* depletion on replication, we used the DNA combing technique and measured replication-fork velocity at the single-molecule level. Analysis of over 250 replication forks in wild-type and knockdown cells confirmed that *NORAD* and RBMX depletion significantly reduced replication-fork velocity (Fig. 4a); the observed effect-size was comparable to previously published TOP1 knockdown data[18]. Thus, *NORAD* may directly or indirectly affect DNA replication even in the absence of additional DNA damage stimulus.

We tested whether *NORAD* depletion also affects cell-cycle progression. We labelled newly synthesized DNA with 5-ethynyl-2′-deoxyuridine (EdU) and measured EdU incorporation and total DNA content by fluorescence-activated cell sorting. We observed a clear decrease in S phase accompanied by increased G1 phase in *NORAD*-, RBMX- and TOP1-depleted cells (Fig. 4b, c and Extended Data Fig. 5b–f). Consistent with these findings, impaired replication-fork progression has been linked to chromosome mis-segregation[28] (as observed in *NORAD* and RBMX knockdown cells), which in turn can trigger a cell-cycle arrest in the subsequent G1 phase[29]. Importantly, a G1 arrest alone cannot explain the reduction in replication-fork velocity observed above. We next examined whether the *NORAD*–RBMX interaction is important for this effect on cell-cycle progression. Expression of full-length *NORAD* in *trans* was sufficient to rescue cell-cycle defects in *NORAD*-knockdown cells (Fig. 4b and Extended Data Fig. 5b, e). By contrast, a *NORAD* construct that lacks the RBMX-binding site decreased S phase and increased G2/M phase; this contrasts with *NORAD* knockdown and may point towards an altered molecular function of truncated *NORAD* (Fig. 4b and Extended Data Fig. 5b, e). Deletion of the RBMX-binding site may therefore act as a dominant negative alteration, which indicates that the RBMX-binding region is required for correct *NORAD* function.

Our results link the known function of members of the NARC1 complex (particularly TOP1) in preventing replication stress and genome instability[16,18,19,27] to the role of *NORAD* in suppressing aneuploidy. Importantly, we demonstrate that the RBMX-binding region in *NORAD* contributes to *NORAD* function, presumably by promoting NARC1 assembly.

It has widely been suggested that lncRNAs participate in assembling groups of proteins, but lncRNA–protein complexes have been fully characterized for only a few lncRNAs; these include *XIST*[2], *TERC*[3], *NEAT1*[4], *MALAT1*[5] and *HOTAIR*[6]. Our results demonstrate that *NORAD* is essential for the assembly of the ribonucleoprotein complex NARC1, which physically links proteins known to be involved in DNA replication or repair but not known to act together. We suggest a model in which depletion of *NORAD* or deletion of its RBMX-binding site disrupts NARC1, which alters replication-fork velocity and impairs

cell-cycle progression. It is tempting to speculate that altered DNA replication causes cells to accumulate the observed chromosome segregation defects, a known cause of genomic instability and aneuploidy[28,29] (Fig. 4d). While our data demonstrates a central role of NARC1 in the *NORAD* phenotype, other proteins or complexes may contribute to additional aspects of *NORAD* function.

The precise mechanism or mechanisms by which *NORAD* promotes NARC1 assembly remain to be elucidated but might include (1) inducing a conformational change in RBMX, (2) recruiting a large number of RBMX molecules to its 5′ end to create a protein interaction scaffold or (3) using other direct interactions to bring NARC1 members into close proximity. RBMX encodes a large low-complexity domain that can self-assemble and undergo phase separation in vitro[30]. Binding of RBMX to *NORAD* may nucleate the formation of higher-order RBMX assemblies that facilitate binding of other proteins that contain a low-complexity domain.

In addition to these structural features, *NORAD* has unusual functional features in that *NORAD* localization to the nucleus can be triggered by DNA damage, which may allow cells to rapidly assemble NARC1 or to re-localize pre-assembled complexes without the need for additional protein synthesis.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, statements of data availability and associated accession codes are available at https://doi.org/10.1038/s41586-018-0453-z.

Received: 10 January 2018; Accepted: 17 July 2018;
Published online 27 August 2018.

1. Guttman, M. et al. Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in mammals. *Nature* **458**, 223–227 (2009).
2. McHugh, C. A. et al. The *Xist* lncRNA interacts directly with SHARP to silence transcription through HDAC3. *Nature* **521**, 232–236 (2015).
3. Zappulla, D. C. & Cech, T. R. Yeast telomerase RNA: a flexible scaffold for protein subunits. *Proc. Natl Acad. Sci. USA* **101**, 10024–10029 (2004).
4. Clemson, C. M. et al. An architectural role for a nuclear noncoding RNA: *NEAT1* RNA is essential for the structure of paraspeckles. *Mol. Cell* **33**, 717–726 (2009).
5. Tripathi, V. et al. The nuclear-retained noncoding RNA MALAT1 regulates alternative splicing by modulating SR splicing factor phosphorylation. *Mol. Cell* **39**, 925–938 (2010).
6. Tsai, M.-C. et al. Long noncoding RNA as modular scaffold of histone modification complexes. *Science* **329**, 689–693 (2010).
7. Kopp, F. & Mendell, J. T. Functional classification and experimental dissection of long noncoding RNAs. *Cell* **172**, 393–407 (2018).
8. Lee, S. et al. Noncoding RNA *NORAD* regulates genomic stability by sequestering PUMILIO proteins. *Cell* **164**, 69–80 (2016).
9. Tichon, A. et al. A conserved abundant cytoplasmic long noncoding RNA modulates repression by Pumilio proteins in human cells. *Nat. Commun.* **7**, 12209 (2016).
10. Baltz, A. G. et al. The mRNA-bound proteome and its global occupancy profile on protein-coding transcripts. *Mol. Cell* **46**, 674–690 (2012).
11. Esakova, O. & Krasilnikov, A. S. Of proteins and RNA: the RNase P/MRP family. *RNA* **16**, 1725–1747 (2010).
12. Wang, T. et al. Identification and characterization of essential genes in the human genome. *Science* **350**, 1096–1101 (2015).
13. Adamson, B., Smogorzewska, A., Sigoillot, F. D., King, R. W. & Elledge, S. J. A genome-wide homologous recombination screen identifies the RNA-binding protein RBMX as a component of the DNA-damage response. *Nat. Cell Biol.* **14**, 318–328 (2012).
14. Matsunaga, S. et al. RBMX: a regulator for maintenance and centromeric protection of sister chromatid cohesion. *Cell Reports* **1**, 299–308 (2012).
15. Pommier, Y., Sun, Y., Huang, S. N. & Nitiss, J. L. Roles of eukaryotic topoisomerases in transcription, replication and genomic stability. *Nat. Rev. Mol. Cell Biol.* **17**, 703–721 (2016).
16. Mahajan, K. hPso4/hPrp19: a critical component of DNA repair and DNA damage checkpoint complexes. *Oncogene* **35**, 2279–2286 (2016).
17. García-Muse, T. & Aguilera, A. Transcription-replication conflicts: how they occur and how they are resolved. *Nat. Rev. Mol. Cell Biol.* **17**, 553–563 (2016).
18. Tuduri, S. et al. Topoisomerase I suppresses genomic instability by preventing interference between replication and transcription. *Nat. Cell Biol.* **11**, 1315–1324 (2009).
19. Gan, W. et al. R-loop-mediated genomic instability is caused by impairment of replication fork progression. *Genes Dev.* **25**, 2041–2056 (2011).
20. Zhang, N., Kaur, R., Akhter, S. & Legerski, R. J. Cdc5L interacts with ATR and is required for the S-phase cell-cycle checkpoint. *EMBO Rep.* **10**, 1029–1035 (2009).
21. Maréchal, A. et al. PRP19 transforms into a sensor of RPA-ssDNA after DNA damage and drives ATR activation via a ubiquitin-mediated circuitry. *Mol. Cell* **53**, 235–246 (2014).
22. Jeronimo, C. et al. Systematic analysis of the protein interaction network for the human transcription machinery reveals the identity of the 7SK capping enzyme. *Mol. Cell* **27**, 262–274 (2007).
23. Boeing, S. et al. Multiomic analysis of the UV-induced DNA damage response. *Cell Reports* **15**, 1597–1610 (2016).
24. Domínguez-Sánchez, M. S., Barroso, S., Gómez-González, B., Luna, R. & Aguilera, A. Genome instability and transcription elongation impairment in human cells depleted of THO/TREX. *PLoS Genet.* **7**, e1002386 (2011).
25. Li, M., Pokharel, S., Wang, J.-T., Xu, X. & Liu, Y. RECQ5-dependent SUMOylation of DNA topoisomerase I prevents transcription-associated genome instability. *Nat. Commun.* **6**, 6720 (2015).
26. Abbas, M., Shanmugam, I., Bsaili, M., Hromas, R. & Shaheen, M. The role of the human psoralen 4 (hPso4) protein complex in replication stress and homologous recombination. *J. Biol. Chem.* **289**, 14009–14019 (2014).
27. Miao, Z.-H. et al. Nonclassic functions of human topoisomerase I: genome-wide and pharmacologic analyses. *Cancer Res.* **67**, 8752–8761 (2007).
28. Burrell, R. A. et al. Replication stress links structural and numerical cancer chromosomal instability. *Nature* **494**, 492–496 (2013).
29. Janssen, A., van der Burg, M., Szuhai, K., Kops, G. J. P. L. & Medema, R. H. Chromosome segregation errors as a cause of DNA damage and structural chromosome aberrations. *Science* **333**, 1895–1898 (2011).
30. Liu, N. et al. N6-methyladenosine alters RNA structure to regulate binding of a low-complexity protein. *Nucleic Acids Res.* **45**, 6051–6063 (2017).

Acknowledgements We thank C. A. McHugh for help setting up RAP MS; K. A. Knouse for help imaging mitotic cells; C. W. Garvie for help with size-exclusion chromatography; L. Ludwig for help with cell-cycle data visualization; L. Gaffney for artwork; T. Wang, B. Cleary, S. R. Grossman, M. Yassour, C. M. Vockley, B. Cimini, K. W. Karhohs, M. Doan, S. A. Myers, D. R. Mani and V. G. Sankaran for discussions. M.M. is supported by a Deutsche Forschungsgemeinschaft (DFG) Research Fellowship. J.M.E. is supported by the Harvard Society of Fellows. M.G. is supported by an NIH Director's Early Independence Award (DP5OD012190), the NIH 4DN program Nucleome Project (U01 DA040612 and U01 HL130007), the New York Stem Cell Foundation, the Edward Mallinckrodt Foundation, Sontag Foundation, Searle Scholars Program, Pew-Steward Scholars program and funds from the California Institute of Technology. M.G. is a NYSCF-Robertson Investigator. Work in the Lander Laboratory is supported by the Broad Institute.

Reviewer information *Nature* thanks R. Bonasio, S. Diedrichs and the other anonymous reviewer(s) for their contribution to the peer review of this work.

Author contributions M.M. and E.S.L. conceived and designed the study. M.M. performed and analysed all experiments; C.T.N. and K.S. assisted with several experiments, including CLIP. V.S. helped with RNA FISH experiments. C.T.N. analysed RNA FISH and PLA images. C.R.H. collected mass spectrometry data under supervision of M.S. and S.A.C. and helped with design of mass spectrometry experiments. M.G. helped setting up RAP MS, L.H. and J.M.E. developed computational tools and analysed CLIP data. J.M.E. and C.P.F. helped with CRISPR interference and contributed analytical ideas. J.C. performed evolutionary analysis. J.C.U. performed differential gene expression and alternative splicing analysis. M.M. and E.S.L. wrote the manuscript with input from all authors. E.S.L. supervised the work and obtained funding.

Competing interests The Broad Institute holds patents and has filed patent applications on technologies related to other aspects of CRISPR.

Additional information
Extended data is available for this paper at https://doi.org/10.1038/s41586-018-0453-z.
Supplementary information is available for this paper at https://doi.org/10.1038/s41586-018-0453-z.
Reprints and permissions information is available at http://www.nature.com/reprints.
Correspondence and requests for materials should be addressed to M.M. or E.S.L.
Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

METHODS

No statistical methods were used to predetermine sample size. The experiments were not randomized and investigators were not blinded to allocation during experiments and outcome assessment.

Tissue culture. We maintained HCT116 cells (ATCC) in McCoy's 5A (Thermo Fisher Scientific) with 10% heat-inactivated FBS (HIFBS, Thermo Fisher Scientific), 1 mM sodium pyruvate, 2mM L-glutamine, and 100 units/ml streptomycin and 100 mg/ml penicillin. Cells were grown at 37 °C and 5% CO_2 atmosphere.

Lentivirus production. We plated 700,000 HEK293T cells in 6-well tissue culture plates and grew them for 24 h before transfecting with 1 μg dVPR, 300 ng VSVG, and 1.2 μg transfer plasmid using FuGene HD (Promega). Sixteen hours after transfection we changed the medium to DMEM with 20% HIFBS. At 48 h post-transfection, we collected viral supernatants and filtered them through a 0.45 μM syringe filter before use.

Generation of CRISPR interference cell lines. We generated inducible CRISPR interference (CRISPRi) cell lines by transducing HCT116 cells with a construct expressing rtTA linked by IRES to a neomycin resistance cassette expressed from an EF1α promoter (ClonTech) and selecting with 200 μg/ml G418 (Thermo Fisher Scientific). Next, rtTA-expressing HCT116 cells were transduced with a previously described KRAB–dCas9 construct linked by IRES to BFP[31]. We selected for cells expressing BFP by fluorescence-activated cell sorting. Inducible *NORAD*, RBMX and PUM2 knockdown cell lines were generated by transducing stable CRISPRi lines with sgRNAs (expressed from a previously described sgOpti backbone[31]) and selecting with 1 μg/ml puromycin.

RAP MS. To capture endogenous *NORAD* transcripts, we designed and synthesized 5′ biotinylated 90-mer DNA oligonucleotides (Integrated DNA Technologies) antisense to the target RNA sequence. We used 26 probes that covered the entire *NORAD* sequence, with the exception of regions that matched to other transcripts or genomic regions as previously described[32]. For *NORAD* and *RMRP* antisense purifications we grew 500 million HCT116 cells per RNA. We supplemented cell culture medium with a final concentration of 200 μM 4-thiouridine and grew cells for 8 h before crosslinking. Cells were washed once with PBS and then crosslinked on ice using 0.8 J/cm^2 of 365-nm ultraviolet light in a Stratalinker (Stratagene). Cells were then scraped from culture dishes, washed once with PBS, pelleted by centrifugation at 500*g* for 5 min and flash-frozen in liquid nitrogen for storage at −80 °C. Preparation of total cell lysates was performed as previously described[2]. For antisense purification of crosslinked protein–RNA complexes we included the following modifications to the previously described procedure: all buffers were pre-heated to 55 °C. We used 50 μg pooled antisense probes for 500 million lysed cells. For pre-clear of lysates and capture of RNA/DNA hybrids we used 5 ml streptavidin Dynabeads MyOne C1 magnetic beads (Thermo Fisher Scientific) for 500 million cells. Elution of captured proteins from streptavidin beads was achieved by digesting nucleic acids using 250 U of benzonase (Millipore), 25 U RNase A and 1000 U RNase T1 (Thermo Fisher Scientific) for 8 h at 37 °C. Trichloroacetic acid-precipitated proteins were reconstituted in 8 M urea and 50 mM Tris-HCl pH 7.8 and stored at −20 °C until processing for mass spectrometry.

Protein digestion for RAP MS. RAP-captured proteins were resuspended in 40 μl of digestion buffer (8 M urea, 50 mM Tris-HCl pH 7.8), reduced (1 μl of 160 mM DTT, 30 min, room temperature) and alkylated (1.6 μl of 250 mM IAA, 45 min, dark, room temperature), followed by a 2 h Lys-c digestion (0.1 μg per sample) at room temperature. Next, the samples were diluted with 120 μl of 100 mM Tris-HCl pH 7.8 to a final concentration of <2 M urea, and 0.5 μg of trypsin was added for overnight digestion at room temperature with agitation. Samples were quenched with 8.5 μl of formic acid and desalted on 4-punch STAGE-Tips as previously described[33].

iTRAQ labelling of peptides and BRP fractionation for RAP MS. Desalted peptides were labelled with iTRAQ4[34] reagent according to the manufacturer's instructions (AB Sciex). Peptides were dissolved in 30 μl of 50 mM triethylamonium bicarbonate (TEAB) pH 8.5 and labelling reagent was added in 70 μl of ethanol. Samples were incubated with labelling reagent for 1 h with agitation, and the reaction was quenched with 5 μl of 1 M Tris-HCl pH 7.8. Differentially labelled peptides were subsequently mixed and prepared for BRP fractionation on 50 mg SepPak columns according to the following protocol: cartridges were prepared for desalting by equilibrating with methanol, 50% acetonitrile (ACN), 1% formic acid and 3 washes with 0.1% TFA. Samples were loaded on to the cartridge and washed 3 times with 1% formic acid. A pH switch was performed before the collection of BRP fractions with 5 mM ammonium formate at pH 10. BRP fractions were collected at the following ACN concentrations: 10% ACN in 5 mM ammonium formate; 15% ACN in 5 mM ammonium formate; 20% ACN in 5 mM ammonium formate; 30% ACN in 5 mM ammonium formate; 40% ACN in 5 mM ammonium formate; and 50% ACN in 5 mM ammonium formate.

Co-immunoprecipitation and MS. To capture RBMX-interacting proteins, we grew 15 million inducible CRISPRi cells with stably integrated *NORAD* sgRNAs for each immunoprecipitation experiment. For *NORAD* depletion samples, we induced knockdown by supplementing cell culture medium with 0.5 μg/ml doxycycline for 72 h, while *NORAD* wild-type samples were grown without doxycycline. Cells were washed in PBS, trypsinized and collected by centrifugation. Cell pellets were washed twice with ice-cold PBS and cell numbers were counted and normalized between knockdown and wild-type samples. Fresh cell pellets containing 15 million cells were lysed in 375 μl co-IP lysis buffer (50 mM Tris-HCl pH 7.5, 150 mM NaCl, 1% NP40, 0.1% sodium deoxycholate, and Halt Protease and Phosphatase Inhibitor Cocktail (Thermo Fisher Scientific)). Lysates were incubated on ice for 30 min and mixed by pipetting every 5–10 min to enhance nuclear lysis. Lysates were cleared by centrifugation at 14,000*g* for 10 min at 4 °C and insoluble material was removed. We pre-cleared lysates by incubating with 50 μl protein A magnetic beads (Thermo Fisher Scientific) for 30 min at 4 °C. Meanwhile, 900 ng RBMX antibody (Cell Signaling #14794) was pre-coupled to 50 μl protein A beads for 45 min at room temperature. We determined the total protein concentration in pre-cleared lysates by BCA assay in triplicates and normalized all samples to contain exactly 2.5 mg total protein. To non-specifically digest all DNA and RNA, we added 50 U benzonase and 1 mM $MgCl_2$ to all lysates. Free RBMX antibody was removed from magnetic beads and benzonase-treated lysates were added to beads and incubated overnight at 4 °C. The next day, supernatant was removed and beads were washed twice in 50 mM Tris-HCl pH 7.5, 150 mM NaCl and 0.05% NP40, followed by two washes in 50 mM Tris-HCl pH 7.5 and 150 mM NaCl. After the last wash, beads were overlaid with 10 μl PBS and immediately subjected to sample preparation for mass spectrometry and TMT labelling.

On-bead protein digestion for co-IP MS. Following immunoprecipitation, washed beads were resuspended in 90 μl of digestion buffer (2 M urea, 50 mM Tris-HCl pH 7.8, 2 mM DTT, 0.005 μg/ml sequencing-grade trypsin) and incubated for 1 h with agitation at 700 rpm. The supernatant was removed and placed in a fresh tube. Beads were washed two times with 60 μl of 2 M urea in 150 mM Tris-HCl pH 7.8, and washes were combined with the supernatant. This procedure was repeated twice to ensure complete removal of proteins from the beads. Supernatants were combined and proteins were reduced (3.5 μl of 500 mM DTT, 30 min, room temperature) and alkylated (9 μl of IAA, 45 min, room temperature, dark), before digestion with 4 μg of trypsin overnight at room temperature with agitation. Samples were acidified (1% formic acid) and desalted on Waters 10 mg Oasis HLB cartridges.

TMT labelling of peptides and BRP Fractionation for co-IP MS. Desalted peptides were labelled with TMT6 reagent according to the manufacturer's instructions (Thermo Fisher Scientific). Peptides were dissolved in 25 μl of HEPES pH 8.5 and 0.2 mg of TMT labelling reagent was added to each sample in 10 μl of ACN. Samples were incubated with labelling reagent for 1 h with agitation. Next, the reaction was quenched with 2 μl of 5% hydroxylamine. Differentially labelled peptides were subsequently mixed and prepared for BRP fractionation on 50 mg SepPak columns according to the following protocol: cartridges were prepared for desalting by equilibrating with methanol, 50% ACN, 1% formic acid and 3 washes with 0.1% TFA. Samples were loaded on to the cartridge and washed 3 times with 1% formic acid. A pH switch was performed with 5 mM ammonium formate at pH 10, collected and run as fraction 1. Subsequent fractions were collected at the following ACN concentrations: 10% ACN in 5 mM ammonium formate; 15% ACN in 5 mM ammonium formate; 20% ACN in 5 mM ammonium formate; 30% ACN in 5 mM ammonium formate; 40% ACN in 5 mM ammonium formate; and 50% ACN in 5 mM ammonium formate.

LC–MS/MS Analysis (RAP MS and co-IP MS). Reconstituted peptides were injected onto a capillary column (Picofrit with 10-μm tip opening, 75-μm diameter, New Objective) packed in-house with 20 cm C18 silica material (1.9 μm ReproSil-Pur C18-AQ medium, Dr Maisch GmbH), and separated on an online nanoflow EASY-nLC 1000 UHPLC system (Thermo Fisher Scientific). Columns were heated to 50 °C in column heater sleeves (Phoenix-ST) to reduce back-pressure during the gradient.

RAP MS experiments. Peptides were separated at a flow rate of 200 nl/min with a linear 120-min gradient from 100% solvent A (3% ACN, 0.1% formic acid) to 35% solvent B (90% ACN, 0.1% formic acid) for 82 min, followed by a 3-min linear increase from 35 to 90% B with a 5-min hold at 60% B before increasing to 90% B for 3 min and holding for 20 min, and equilibrating back at 50% B for 10 min to end the gradient.

Co-IP MS experiments. Peptides in each BRP fraction were separated at a flow rate of 200 nl/min over a linear gradient of 100% A to 20% B for 28 min, with a linear increase from 20% B to 60% B for 16 min, and a hold at 90% B for 5 min before returning to 50% B.

Peptides were analysed on an Orbitrap Q Exactive Plus mass spectrometer (Thermo Fisher Scientific) operated in data-dependent mode. Higher-energy collision dissociation tandem mass spectrometry (HCD MS/MS) scans (resolution = 17,500 for iTRAQ and TMT methods) were taken after each MS1 scan (resolution = 70,000) on the top 12 most abundant ions using an AGC target of

3×10^6 ions for MS1 and 5×10^4 ions for MS2. The isolation widths for MS/MS ions were 1.6 for iTRAQ and TMT methods. The maximum ion fill-time for MS/MS scans was 120 ms, the HCD-normalized collision energy was 29; dynamic exclusion time was set to 20 s, and peptide match and isotope exclusion functions were enabled.

Quantification and identification of peptides and proteins (RAP MS and co-IP MS). All mass spectra were processed using the Spectrum Mill software package v.6.01 pre-release (Agilent Technologies), which includes modules developed for iTRAQ and TMT6-based quantification. Precursor ion quantification was done using extracted ion chromatograms for each precursor ion. The peak area for the extracted ion chromatogram of each precursor ion subjected to MS/MS was calculated in the intervening high-resolution MS1 scans of the LC–MS/MS runs using narrow windows around each individual member of the isotope cluster. Peak widths in both time and m/z domains were dynamically determined on the basis of mass spectrometry scan resolution, precursor charge and m/z, subject to quality metrics on the relative distribution of the peaks in the isotope cluster versus theoretical. Similar MS/MS spectra acquired on the same precursor m/z in the same dissociation mode with ± 60 s were merged. MS/MS spectra with precursor charge >7 and poor quality MS/MS spectra, which failed the quality filter by having a sequence tag length less than 1, were excluded from searching.

For peptide identification, MS/MS spectra were searched against the human Uniprot database to which a set of common laboratory contaminant proteins was appended. Search parameters included: ESI-QEXACTIVE-HCD scoring parameters, trypsin or Lys-c/trypsin enzyme specificity with a maximum of 2 missed cleavage, 40% minimum matched peak intensity, ± 20 ppm precursor mass tolerance, ± 20 ppm product mass tolerance, and carbamidomethylation of cysteins and isobaric labelling of lysines and N-termini as fixed modifications in the RAP MS (iTRAQ) and the immunoprecipitation mass spectrometry (TMT6) experiments with no fixed modification on lysines or N-termini for the size-exclusion chromatography experiment. Oxidation of methionine, N-terminal acetylation and deamidated (N) were allowed as variable modifications, with a precursor MH^+ shift range from -18 to 64 Da. Identities interpreted for individual spectra were automatically designated as valid by optimizing score and delta rank1–rank2 score thresholds separately for each precursor charge state in each LC–MS/MS run, while allowing a maximum target-decoy-based false-discovery rate (FDR) of 1.0% at the spectrum level.

In calculating scores at the protein level and reporting the identified proteins, redundancy is addressed in the following manner: the protein score is the sum of the scores of distinct peptides. A distinct peptide is the single highest scoring instance of a peptide detected through an MS/MS spectrum. MS/MS spectra for a particular peptide may have been recorded multiple times (that is, from different precursor charge states, isolated from adjacent BRP fractions or modified by oxidation of Met), but are still counted as a single distinct peptide. When a peptide sequence over eight residues long is contained in multiple protein entries in the sequence database, the proteins are grouped together and the highest scoring one and its accession number are reported. In some cases in which the protein sequences are grouped in this manner, there are distinct peptides that uniquely represent a lower scoring member of the group (isoforms or family members). Each of these instances spawns a subgroup, and multiple subgroups are reported and counted towards the total number of proteins identified. iTRAQ and TMT ratios were obtained from the protein comparisons export table in Spectrum Mill. To obtain iTRAQ or TMT protein ratios, the median was calculated over all of the distinct peptides assigned to a protein subgroup in each replicate. For RAP MS we required each protein to be detected with more than two unique peptides. To enable precise quantification, we limited our analysis to peptides that are uniquely assigned to a specific protein isoform or family member. For co-IP MS we required more than four unique peptides. For statistical analysis, we used the Limma package[35] in R (https://www.r-project.org/) to calculate multiple comparison adjusted P values using a moderated t-test.

Native RBMX co-IP and size-excusion chromatograpy. To capture native RBMX complexes, we generated stable HCT116 cell lines that express Flag–RBMX–V5. For co-IP and size-exclusion chromatography experiments, we grew 200 million cells. Cells were collected by scraping culture dishes, washed once with PBS and pelleted by centrifugation at 500*g* for 5 min. Fresh cell pellets were lysed in 8 ml co-IP lysis buffer (50 mM Tris-HCl pH 7.5, 150 mM NaCl, 1% NP40, 0.1% sodium deoxycholate, and Halt Protease and Phosphatase Inhibitor Cocktail (Thermo Fisher Scientific)). Lysates were incubated on ice for 30 min and mixed by pipetting every 5–10 min to enhance nuclear lysis. Lysates were cleared by centrifugation at 14,000*g* for 10 min at 4 °C and insoluble material was removed. We pre-cleared lysates by incubating with 2.5 ml protein G magnetic beads (Thermo Fisher Scientific) for 30 min at 4 °C. Meanwhile, 600 μg Flag M2 antibody (Sigma # F1804) was pre-coupled to 2.5 ml protein G beads for 45 min at room temperature. To non-specifically digest DNA and RNA, we added 500 U benzonase to the cell lysate. Free Flag M2 antibody was removed from magnetic beads and benzonase-treated lysates were added to beads and incubated overnight at 4 °C. The next day, supernatant was removed and beads were washed twice in 50 mM Tris-HCl pH 7.5, 150 mM NaCl and 0.05% NP40, followed by two washes in 50 mM Tris-HCl pH 7.5 and 150 mM NaCl. After the last wash, protein complexes were eluted using 250 μg Flag-peptide in 500 μl of 150 mM NaCl, 25 mM Tris pH 7.5, 0.05% IGEPAL. Elutions were incubated 1 h at 4 °C with agitation. Eluates were separated from beads and filtered using a 0.2-μm membrane filter. Size-exclusion chromatography of the RBMX complex was performed using a Superose 6 Increase 10/300 column (GE Healthcare) equilibrated in 150 mM NaCl, 25 mM Tris pH 7.5, 0.05% IGEPAL. We injected 400 μl of the eluate onto the column at a flow rate of 0.4 ml/min and collected 0.5-ml fractions. Two hundred and fifty microliters of each fraction was subjected to trichloroacetic acid-precipitation to concentrate proteins. Protein content was analysed by western blotting and mass spectrometry.

For mass spectrometry analysis, proteins were reduced, alkylated and denatured at 90 °C for 5 min, spun down and loaded separated by SDS–PAGE. The gel was run in 1× MES SDS–PAGE running buffer at 175 V for 40 min, after which it was stained for 2 h in SimplyBlue Safe Stain (Thermo Fisher Scientific) and destained in water overnight. The gel lane was cut into 4 fractions, diced and destained with 50% ACN, 50% 100mM ammonium bicarbonate. Destaining buffer was removed and gel pieces were dehydrated with 300 μl of ACN. ACN was aspirated once the gel pieces were white. One-and-a-half micrograms of trypsin was added to each of the 4 fractions in 100 μl of 100 mM ammonium bicarbonate (pH 8) and incubated overnight at 37 °C. The supernatant from each fraction was collected into in a fresh tube, and the peptides were extracted from the gel pieces by washing twice with 60% ACN, 0.1% formic acid and collecting the extract in the tube with the initial supernatant. Finally, gel pieces were dehydrated with ACN, which was collected with the rest of the extract. Fractions were then dried using a Speedvac concentrator, reconstituted in 3% ACN and 0.1% formic acid, and desalted on C18 Stage Tips[33]. Eluate from each fraction was transferred to HPLC vials, dried down and reconstituted in 5 μl of 3% ACN, 5% formic acid and run on an EasyNLC 1200 coupled to an Orbitrap Q Exactive Plus mass spectrometer. The previously described method for co-IP MS experiments (see 'Co-IP MS experiments' above) was used for analysis, with the only difference being a normalized collision energy of 25, which is routinely used for label-free peptide analysis.

To extract RNA from size-exclusion chromatography fractions containing the protein complex as well as control fractions we Trizol-extracted the remaining 250 μl of sample, and isolated RNA using Direct-zol columns (Zymo Research). We removed rRNA with the NEBNext rRNA Depletion Kit (New England Biolabs) by following the manufacturer's instructions. Finally, we constructed RNA-sequencing libraries using the SMARTer smRNA-Seq Kit (Clontech) by following the manufacturer's instructions. Libraries were sequenced on an Illumina HiSeq 2500 instrument to an average read depth of 15–20 million reads with 50-bp read 1 and 60-bp read 2. We trimmed 5 bp from the beginning of read 1 and 15 bp from the beginning of read 2 before mapping. Reads aligning to rRNA were removed from downstream analysis[36]. Reads were then mapped to hg19 using Bowtie2. Mapping results were restricted to the single best alignment found for any given read. Discordant alignments of paired-end reads were excluded from analysis. Data normalization was performed by scaling coverage values by (1,000,000/total mapped read count).

CLIP. The CLIP protocol below is extensively based on three previously published CLIP methods: irCLIP[37], PAR-CLIP[38] and eCLIP[39].

We constructed the pre-adenylated irCLIP adaptor as previously described[37]. All other oligonucleotides were synthesized as described in the irCLIP protocol, with the exception of reverse transcription primers. We replaced ethyleneglycol spacers with three deoxyuridines and modified the 5′ end of reverse transcription primers to reflect the nucleotide preference of CircLigase II (general structure: /5phos/RNNNNN-6nt-barcode-NNNNNTACCCTTCGCTTCACACACAAG/ideoxyU//ideoxyU//ideoxyU/TACTGAACCGC).

For each CLIP experiment we grew 20 million HCT116 cells in medium supplemented with 200 μM 4-thiouridine for 8 h. Cells were washed once with PBS and then crosslinked on ice using 0.2 J/cm^2 of 365-nm ultraviolet light in a Stratalinker. Cells were then scraped from culture dishes, washed once with PBS, pelleted by centrifugation at 500*g* for 5 min and flash-frozen in liquid nitrogen for storage at −80 °C. To prepare cell lysates, pellets were thawed on ice and resuspended in NP40 lysis buffer (50 mM HEPES pH 7.5, 150 mM KCl, 2 mM EDTA, 1% (v/v) NP40, 0.25 mM DTT, complete EDTA-free protease inhibitor cocktail) and incubated on ice for 10 min. We sonicated cell lysates using a Branson Digital Sonifier with a microtip set at 5 W power for a total of 1 min 30 s in intermittent pulses (0.7-s on, 2.3-s off), followed by RNase I (Thermo Fisher Scientific) digestion (0.5 U/μl, 10 min at 23 °C). Subsequently, we added 15 μl/ml Murine RNase Inhibitor (New England Biolabs), followed by DNA digestion (20 U TURBO DNase (2 U/μl; Thermo Fisher Scientific), 2.5 mM $MgCl_2$ and 0.5 mM CaCl) for 20 min at 37 °C. We incubated samples on ice for 10 min before clearing lysates by centrifugation at

15,000*g* for 15 min. Insoluble material was removed and total protein concentration was determined by BCA assay. Cell lysates were flash-frozen and stored in batches of 10 mg total protein at −80 °C.

For each immunoprecipitation experiment, lysates (10 mg total protein) were thawed on ice and pre-cleared by incubating with protein A/G magnetic beads (using 30 µl/mg total protein) for 30 min at 4 °C. In the meantime, antibodies (6 µg/mg total protein) were coupled to protein A/G magnetic beads (using 30 µl/mg total protein) at room temperature for 45 min (antibodies used: RBMX, Cell Signaling #14794; ALYREF, Bethyl # A302-892A; PUM1, Bethyl # A302-577A; V5, Abcam # ab27671). We removed unbound antibody and added the pre-cleared lysates to antibody-coupled beads and incubated overnight at 4 °C. The following day, we washed the beads 3 times in IP wash buffer (50 mM HEPES pH 7.5, 300 mM KCl, 0.5% (v/v) NP40, 0.25 mM DTT, complete EDTA-free protease inhibitor cocktail), followed by one wash in FastAP buffer (10 mM Tris-HCl pH 8.0, 5 mM $MgCl_2$, 100 mM KCl, 0.02% Triton X-100). Immunopurified protein–RNA complexes were dephosphorylated by resuspending beads in 25 µl FastAP mix (18.5 µl H_2O, 2.5 µl 10× FastAP buffer (Thermo Fisher Scientific), 2.5 U FastAP enzyme (1 U/µl; Thermo Fisher Scientific), 0.5 µl Murine RNase Inhibitor (New England Biolabs)) and incubating for 20 min at 37 °C. In the meantime, we prepared polynucleotide kinase mix (56 µl H_2O, 10 µl 10× PNK buffer (New England Biolabs), 1 µl Murine RNase Inhibitor, 7 µl T4 PNK (10 U/µl; New England Biolabs), 1 µl TURBO DNase) and added 75 µl to each 25 µl sample and incubated 20 min at 37 °C. Beads were separated on a magnet and dephosphorylation reaction was removed before washing beads once in RNA ligation buffer without DTT (50mM Tris-HCl pH 7.5, 10mM $MgCl_2$). Next, 3′ ligation was performed by resuspending beads in 20 µl ligation mix (3 µl H_2O, 2 µl 10× T4 RNA ligation buffer (New England Biolabs), 1 µl DMSO, 1 µl RNase inhibitor, 15 pmoles pre-adenylated 3′ adaptor, 10 µl 50% PEG 8000, 2 µl T4 RNA Ligase 1 High Concentration (New England Biolabs)) using low-retention pipette tips and incubated overnight at 16 °C with agitation. The next day we added 7 µl 4× NuPAGE LDS Sample Buffer (Thermo Fisher Scientific) to ligation reactions and incubated samples for 10 min at 75 °C. Protein–RNA complexes were resolved by SDS–PAGE using NuPAGE 4–12% Bis-Tris-HCl Gels (Thermo Fisher Scientific) at 200 V for 1 h, followed by transfer to a nitrocellulose membrane using the iBlot Dry Blotting System (Thermo Fisher Scientific). Protein–RNA complexes were visualized using the Odyssey Clx infrared imager (LI-COR) and desired complexes were excised from membrane using a clean scalpel. Membrane pieces were immediately subjected to proteinase K treatment by adding 250 µl proteinase K solution (4 mg/ml Proteinase K (New England Biolabs), 100 mM Tris-HCl pH 7.5, 150 mM NaCl, 12.5 mM EDTA, 1% (w/v) SDS) and incubating 1 h at 55 °C. Following proteinase K treatment, RNA was phenol-chloroform extracted using Heavy Phase Lock Gel tubes (5Prime) and purified with the Zymo RNA Clean & Concentrator-5 kit by following the manufacturer's instructions for small and large RNAs. We eluted RNA in 7 µl H_2O and combined it with 10 pmoles of reverse transcription primer. Samples were heated to 72 °C for 2 min and snap-cooled on ice. Reverse transcription was performed with the SuperScript III kit (Thermo Fisher Scientific) by combining RNA samples with 4 µl 5× First Strand Buffer, 2 µl 0.1M DTT and 6 µl 2mM dNTPs. Samples were incubated at 50 °C for 3 min before adding 1 µl SuperScript III reverse transcription enzyme (200 U/µl) and incubating 1 h at 42 °C. For each reverse transcription reaction, we washed 5 µl MyOne streptavidin C1 beads twice in NT2 buffer (50 mM Tris-HCl pH 7.5, 150 mM NaCl, 1 mM $MgCl_2$ and 0.0005% NP40) and then resuspended the beads in 100 µl NT2 buffer. The resuspended beads were then added to the reverse transcription reaction and the mixture was subsequently incubated for 15 min at room temperature. Beads were washed twice in streptavidin wash buffer (20 mM Tris-HCl pH 7.5, 120 mM NaCl, 2 5mM KCl, 5 mM EDTA, 1% Triton X-100, 1% sodium deoxycholate) and twice in PBS to remove unbound reverse transcription primer. Finally, we resuspended beads in 10 µl of freshly prepared elution buffer (8.25 µl H2O, 1 µl 1 µM elution oligonucleotide (CTGAACCGCTCTTCCGATCT), 0.75 µl of 50 mM $MnCl_2$) and heated the reaction for 5 min at 95 °C, 1 min at 75 °C, followed by a ramp of 0.1 °C per s to 60 °C and holding at 60 °C for 15 min. Once the 60-°C incubation temperature was reached, we prepared CircLigase mix (2 µl H_2O, 2 µl 10× CircLigase-II buffer (Epicentre), 0.25 µl 50mM $MnCl_2$, 4 µl 5M betaine, 1 µl CircLigase-II (Epicentre)) and added 10 µl to each elution without removing beads. We incubated the reaction 2 h at 60 °C. Following incubation, we added 2 reaction volumes Agencourt AMPure XP beads (Beckman Coulter) and 5 reaction volumes isopropanol and incubated 15 min at room temperature. Supernatant was removed and beads were washed once with 80% ethanol (v/v), air-dried and eluted by resuspending dry beads in 25 µl H_2O, heating for 2 min at 95 °C and isolating supernatants. We used 12 µl cDNA for PCR amplification (program: 2 min at 98 °C followed by 12–15 cycles of 98 °C for 15 s, 65 °C for 30 s and 72 °C for 30 s, followed by a final 20-min extension at 72 °C) in a 50 µl reaction using 25 µl 2× NEBNext Q5 Hot Start HiFi PCR Master Mix (New England Biolabs), 12 µl H_2O and 1 µl of 25µM PCR1 primer mix containing P3_PCR1 (GCATTCCTGCTGAACCGCTCTTCCGATCT) and P6_PCR1 (TTTCCCCTTGTGTGTGAAGCGAAGGGTA) primers. PCR reactions were subjected to two consecutive rounds of purification using 1.5 volumes of Agencourt AMPure XP beads and two 70% ethanol washes. DNA was eluted in 14 µl H_2O and subjected to a second PCR amplification (program: 2 min at 98 °C followed by 3–6 cycles of 98 °C for 15 s and 72 °C for 45 s, followed by a final 2-min extension at 72 °C) in a 50 µl reaction using 12 µl purified PCR1, 25 µl 2× NEBNext Q5 Hot Start HiFi PCR Master Mix (New England Biolabs), 12 µl H_2O and 1 µl of 25 µM PCR2 primer mix containing P3_PCR2 (CAAGCAGAAGACGGCATACGAGATCGGTCTCGGCATTCCTGCTGAACCGCTCTTCCGATCT) and P6_PCR2 (AATGATACGGCGACCACCGAGATCTACACTCTTTCCCCT TGTGTGTGAAGC GAAGGGTA) primers. Upon completion of PCR amplification, we added 10 µl ExoSAP-IT PCR Product Cleanup Reagent (Thermo Fisher Scientific) and incubated reactions for 15 min at 37 °C. Reactions were purified using 1.1 volumes of Agencourt AMPure XP beads and two 70% ethanol washes, followed by elution of air-dried beads in 10 µl H_2O. The concentration of final libraries was determined with the Qubit dsDNA HS Assay (Thermo Fisher Scientific) and library sizes were analysed on a High Sensitivity DNA Bioanalyzer Chip (Agilent).

Size-matched input libraries (SM input)[39] were prepared by resolving 1–2% of input lysates by SDS–PAGE using NuPAGE 4–12% Bis-Tris-HCl Gels (Thermo Fisher Scientific) at 200 V for 1 h. SDS–PAGE gels were transferred to a nitrocellulose membrane using the iBlot Dry Blotting System (Thermo Fisher Scientific) and proteins migrating at the molecular weight range of the target protein were excised using a clean scalpel. RNA was released by proteinase K treatment and purified as described in the previous section. We performed end-repair of input RNA by adjusting RNA volume to 19.5 µl with H_2O and adding 2.5 µl 10× FastAP buffer (Thermo Fisher Scientific), 2.5 U FastAP enzyme (1 U/µl; Thermo Fisher Scientific), 0.5 µl Murine RNase Inhibitor (New England Biolabs)) and incubating for 20 min at 37 °C. In the meantime, we prepared Polynucleotide kinase mix (56 µl H_2O, 10 µl 10× PNK buffer (New England Biolabs), 1 µl Murine RNase Inhibitor, 7 µl T4 PNK (10 U/µl; New England Biolabs), 1 µl TURBO DNase) and added 75 µl to each 25 µl sample and incubated samples for 20 min at 37 °C. RNA was purified with the Zymo RNA Clean & Concentrator-5 kit using the manufacturer's instructions for small and large RNAs. RNA was eluted in 5 µl H_2O and combined with 25 µl ligation mix (3 µl 10× T4 RNA ligation buffer (New England Biolabs), 1.5 µl DMSO, 1.5 µl RNase inhibitor, 15 pmoles pre-adenylated 3′ adaptor, 15 µl 50% PEG 8000, 3 µl T4 RNA Ligase 1 High Concentration (New England Biolabs)) using low-retention pipette tips and incubated for 2 h at 23 °C with agitation. Ligation reactions were purified to remove free 3′ adaptor using two consecutive Silane bead purifications. For each reaction, we washed 15 µl Silane beads (Thermo Fisher Scientific) twice in 1 ml RLT buffer (Qiagen), resuspended beads in 90 µl RLT and combined 90 µl beads in RLT with 30 µl ligation reaction. We added 0.7 volumes 100% ethanol and incubated mixtures 10 min at room temperature. Supernatant was removed and beads were washed twice with 70% ethanol before eluting air-dried beads in 9 µl H_2O. We used 7 µl of the eluted RNA for reverse transcription and proceeded with the library preparation as described in the above section.

Computational analysis of CLIP data. We sequenced CLIP and corresponding SM input libraries on an Illumina HiSeq 2500 to an average read depth of 30–50 million reads with 52-bp read 1 and 35-bp read 2. The first read includes a 6-nt barcode added during reverse transcription (see 'CLIP' above). After processing to separate samples based on inline barcodes, sequencing reads collected from all CLIP experiments were first mapped to hg19 using TopHat (v.2.0.8)[40]. Reads aligning to rRNA were removed from downstream analysis, as previously described[36]. Duplicate reads were identified and removed using Picard's MarkDuplicates program. Peak calling was performed with the MACS2[41] algorithm to identify genomic coordinates where experimental conditions (protein IP) were significantly enriched for reads relative to size-matched controls (SM input). Peak calling was performed without a shifting model and the band width to compute fragment size was set to 100 bp. Significant peaks are reported with FDR correction of $q = 0.05$. Significant peaks were further filtered to include only regions with an average minimum depth of two reads in the size-matched control condition. To compile significant results across replicate experiments, we intersected the intervals from the peak calling output of each replicate. Normalized coverage in the intersection peaks was first calculated separately for each replicate as the average depth at a given peak divided by the total number of reads after correcting for the observed duplication rate. The mean of the relative fold change between the two replicates was calculated for each peak and peaks that did not show a twofold or greater change in both replicates were excluded. We report a CLIP signal score for a given peak as the product of enrichment (average fold change) and the peak length (see Supplementary Table 3).

RNA-sequencing and analysis. We performed RNA-sequencing on cells that stably expressed individual sgRNAs targeting the *NORAD* promoter, with or without doxycycline-induced KRAB–dCas9 expression. We performed at least 2 biological replicate experiments for knockdown and control conditions after 24 h, 48 h and 96 h of KRAB–dCas9 induction. RNA-sequencing libraries were constructed

as previously described[36]. Reads were pseudo-aligned to hg19 (ENSEMBL transcripts) using kallisto[42] with an index of either 31 or 21 k-mers. Estimated counts were collapsed across transcripts into genes and differential expression analysis was performed using DESeq2[43]. Genes with an absolute $\log_2$(fold change) > 1 and FDR < 0.05 were considered as differentially expressed. *P* values for differential gene expression were corrected using the Benjamini–Hochberg procedure to derive an FDR.

Alternative splicing analysis. Percentage spliced in (PSI) for different exons or introns was calculated using SUPPA2[44] based on isoform transcripts per million (TPM) estimates from kallisto for skipping exon, alternative 5′ or 3′ splice sites, mutually exclusive exons, retained intron and alternative first or last exons. Differential PSI was calculated using diffSplice[45] with the parameters '–area 1000–tpm-threshold 5–lower-bound 0.00 -gc'. Events with a change in PSI $> 20\%$ and FDR < 0.05 were considered as differentially used. *P* values across putative splicing events were corrected using the Benjamini–Hochberg procedure to derive an FDR.

RNA extraction and RT–qPCR. We extracted RNA from 20,000–50,000 cells per experiment in RLT buffer (Qiagen) using Dynabeads MyOne Silane beads (Thermo Fisher Scientific), treated samples with TURBO DNase (Thermo Fisher Scientific) and cleaned again with Silane beads. We used AffinityScript reverse transcriptase (Agilent Technologies) and random nonamer primers to convert RNA to cDNA. We performed qPCR using SYBR Green I Master Mix (Roche) and calculated differences using the $\Delta\Delta C_t$ method versus *GAPDH*. To achieve power to detect small effects in gene expression, we performed three technical qPCR replicates (from the same cDNA) and took the median value for further analysis. We used the following RT–qPCR primers in this study. *RBMX* forward primer: CAGTTCGCAGTAGCAGTGGA, *RBMX* reverse primer: TCGAGGTGGACCTCCATAAC; *NORAD* forward primer: CTCTGCTGT GGCTGCCC, *NORAD* reverse primer: GGGTGGGAAAGAGAGGTTCG; *PUM2* forward primer: GGGAGCTTCTCACCATTCAATG, *PUM2* reverse primer: CCA TGAAAACCCTGTCCAGATC; *GAPDH* forward primer: AGCCACATCGC TCAGAC AC, *GAPDH* reverse primer: GCCCAATACGACCAAATCC; *MALAT1* forward primer: AGTTCAGTGTTGGGGCAATC, *MALAT1* reverse primer: GTTCTTCCGCTCAAATCCTG; *TOP1* forward primer: TCGAAGCGG ATTTCCGATTGA, *TOP1* reverse primer: CTTTGTGCCGGTGTTCTCGAT.

Co-IP western blot. Co-IP experiments were carried out as described above (see 'Co-immunoprecipitation and MS'). The following antibodies were used for immunoprecipitation reactions: RBMX, Cell Signaling #14794, or Santa Cruz Biotechnology # sc-14581; ALYREF, Bethyl # A302-892A; TOP1, Bethyl # A302-589A; CDC5L, Bethyl #A301-681A. Following the last washing step, we resuspended beads in 20 μl Pierce IgG Elution Buffer (Thermo Fisher Scientific) and incubated them for 20 min at room temperature with agitation. We collected supernatants and added 7 μl 4× NuPAGE LDS Sample Buffer (Thermo Fisher Scientific), followed by a 3-min incubation at 95 °C. Proteins were resolved by SDS–PAGE using NuPAGE 4–12% Bis-Tris-HCl Gels (Thermo Fisher Scientific) at 200 V for 1 h, followed by transfer to a nitrocellulose membrane using the iBlot Dry Blotting System (Thermo Fisher Scientific). Proteins larger than 150 kDa in size were resolved on NuPAGE 3–8% Tris-Acetate Gels (Thermo Fisher Scientific). Western blots were performed using the iBind Western System (Thermo Fisher Scientific). For protein detection, we used the following primary antibodies: RBMX, Cell Signaling #14794, or Santa Cruz Biotechnology #sc-14581; ALYREF, Santa Cruz Biotechnology #sc-32311; TOP1, Santa Cruz Biotechnology # sc-32736; CDC5L, Santa Cruz Biotechnology #sc-81220. We used the following secondary antibodies: IRDye 680RD Goat anti-Mouse IgG (H + L) (LI-COR), IRDye 800CW Goat anti-Rabbit IgG (H + L) (LI-COR), IRDye 800CW Donkey anti-Goat IgG (H + L) (LI-COR). For visualization of bands, we used the Odyssey Clx infrared imager system (LI-COR).

***NORAD* conservation analysis.** We tested for conservation of *NORAD* transcription across 11 mammalian species: human, chimpanzee, gorilla, orangutan, rhesus macaque, mouse, rat, ferret, dog, cow and armadillo. Because expression of *NORAD* is highest in human brain[46] we checked for transcription in brain tissue from these 11 species. Raw RNA-sequencing read data were downloaded from previous studies[47–49] and mapped to respective genomes using STAR v.2.5.2a[50], with gene annotations from Ensembl Release 91[51] as a reference guide. For each RNA-sequencing library, a de novo transcriptome was made using Stringtie v.1.3.3b[52], using default parameters. Samples from the same species were then merged using the stringtie–merge option. To find reciprocal best hits, we used nucleotide BLAST with default parameters. Multiple sequence alignment was created using MAFFT[53] with gap penalty reduced to 1.0.

Subcellular fractionation. We prepared nuclear and cytoplasmic extracts from freshly grown HCT116 cells using the PARIS Kit (Thermo Fisher Scientific) by following the manufacturer's instructions. We used ~10 million cells for each fractionation experiment and analysed extracts by western blot or RT–qPCR as described in the corresponding sections.

smRNA FISH. smRNA FISH experiments were performed using the ViewRNA Cell Plus Assay Kit (Thermo Fisher Scientific) and following the manufacturer's instructions. We grew 50,000 cells in black 12-well glass-bottom plates (Cellvis) for 24 h. To induce DNA damage, we supplemented culture medium with doxorubicin (1 μM) or camptothecin (200 nM) for 12 h. We washed cells once with PBS before fixation. Cells were then fixed and permeabilized simultaneously in fixation/permeabilization buffer for 30 min at room temperature on a rotating plate. After three brief washes in PBS, we incubated cells with the appropriate probe set diluted 1:100 in probe set diluent for 2 h at 40 °C, then with preamplifier mix at 40 °C for 70 min, followed by amplifier mix at 40 °C for 70 min, and finally label probe mix at 40 °C for 60 min. For nuclei staining, we incubated the cells for 2 min at room temperature with 1× ViewRNA Cell Plus DAPI in PBS. Cells were then washed three times in PBS and then incubated with Alexa Fluor 647 phalliodin (Cell Signaling Technology) diluted 1:20 in PBS for 15 min at room temperature for staining of actin filaments. After a final set of washes, we covered cells with ProLong Gold Anti-Fade Reagent (Cell Signaling Technology) and stored the plates at 4 °C until imaging. The probe sets and corresponding fluorophores were type 1 – *NORAD* and *MALAT1* (Alexa Fluor 546) and type 4 – *GAPDH* (Alexa Fluor 488). Confocal microscopy was performed using a Nikon Eclipse Ti1 with Andor Yokogawa Spinning Disk Revolution WD system.

Quantification of RNA FISH images. For three-dimensional FISH image analysis, *Z*-stacks were exported such that the top and bottom slices were the beginning and end of DAPI signal in the *z* direction. Quantification of FISH foci was done with FISH-quant[54] in MATLAB (version R2017b) following the software's instructions for mature mRNA quantification. Before spot detection, a dual Gaussian filter was applied to the images in FISH-quant using the default settings. The outline of nuclei and cells were determined automatically with the Cell Segmentation Tool in FISH-quant and a modified version of a Cell Profiler pipeline provided in the FISH-quant repository. In Cell Profiler (v.2.2.0)[55], nuclear boundaries were determined by the Otsu method guided by DAPI staining. Identified nuclei were then used as seeds to identify the boundaries of cells by the watershed method aided with the phalloidin stain. For all probes, the local maximum strategy of spot pre-detection was used. Settings for thresholding pre-detected spots were optimized for each probe separately to account for differences in signal intensity.

In situ proximity ligation assay. In situ proximity ligation assay (PLA) was performed using the Duolink PLA platform (Sigma) and following the manufacturer's instructions. Cells were plated in black, glass-bottom 96-well plates the day before the experiment and allowed to grow overnight at 37 °C. Cells were fixed with 4% paraformaldehyde in PBS for 15 min at room temperature, washed three times in PBS and then permeabilized with 0.5% Triton X-100 in PBS for 15 min at room temperature. Cells were blocked for 1 h at 37 °C in a humidified chamber using the Duolink blocking solution, and subsequently stained with primary antibodies diluted 1:250 in Duolink antibody diluent for 1 h at room temperature. DuoLink PLA probes (Rabbit PLUS and Mouse MINUS) were added for 1 h at 37 °C. The ligation and subsequent amplification steps were performed for 30 min and 100 min, respectively, at 37 °C. Upon completion of the assay, cells were overlaid with Duolink mounting medium with DAPI. Two sets of primary antibody pairs were used: rabbit anti-RBMX (Cell Signaling Technologies #14794) was paired with mouse anti-TOP1 (Thermo Fisher Scientific #435900); mouse anti-Flag (Cell Signaling Technologies #8146S) (targeting Flag–RBMX) was paired with rabbit anti-TOP1 (Bethyl #A302-589A).

Quantification of PLA images. For PLA signal quantification we used Cell Profiler 3.0.0. Separate maximum intensity projections for each channel were exported. Nuclei and PLA-signal segmentation was performed using the minimal cross entropy thresholding method. We applied default settings for nuclei segmentation, whereas the PLA signal detection required more stringent thresholding to distinguish individual spots within clusters. A size filter was applied to exclude overlapping nuclei from the analysis. The total nuclear PLA spot count was normalized to the total nuclear area for each cell.

Immunostaining of cultured cells for anaphase nuclei imaging. We induced knockdown in CRISPRi cells with stably integrated sgRNAs by supplementing cell culture medium with 0.5 μg/ml doxycycline for 48 h. Cells were then trypsinized and plated in multi-well glass-bottom plates (Cellvis), again supplementing culture medium with doxycycline in knockdown cells, and grown for an additional 24 h. We removed culture medium, rinsed each well in PBS and fixed cells in 4% paraformaldehyde (PFA) for 10 min at room temperature. All subsequent manipulation steps were carried out in a humidified chamber. PFA was removed, cells were washed twice in PBS and permeabilized by incubating with PBS + 0.1% Triton X-100 for 10 min at room temperature. Following permeabilization, we blocked cells in PBS containing 4% BSA (Roche), 10% goat serum (Sigma Aldrich) and 0.1% Triton X-100 for 30 min at room temperature. Primary antibodies (anti-α-tubulin—FITC antibody, Sigma Aldrich #F2168 (1:1,000); anti-centromere antibodies, Antibodies Incorporated # 15-234-0001 (1:200)) were diluted in blocking buffer (PBS containing 4% BSA (Roche), 10% goat serum (Sigma Aldrich) and

0.1% Triton X-100) and incubated in the dark for 2 h at room temperature or overnight at 4 °C. Following antibody incubation, cells were washed 3 times in PBS + 0.1% Triton X-100 and incubated 10 min at room temperature in between each washing step. Secondary antibody (Goat anti-Human IgG (H+L) Cross-Adsorbed, Alexa Fluor 568, Thermo Fisher Scientific #A-21090 (1:250)) was diluted in blocking buffer and added to cells for 1 h at room temperature in the dark. Cells were washed 3 times in PBS + 0.1% Triton X-100 and incubated 10 min at room temperature in between each washing step. Following removal of any residual washing buffer, we covered cells with ProLong Gold Antifade reagent containing DAPI (Thermo Fisher Scientific) and allowed them to cure overnight in the dark before imaging. Confocal microscopy was performed using a Nikon Eclipse Ti1 with Andor Yokogawa Spinning Disk Revolution WD system.

Generation of *NORAD trans* rescue constructs and cell lines. We synthesized the full-length *NORAD* cDNA (Genewiz) and cloned it into pDONR221 (Thermo Fisher Scientific). Using Gateway technology, we cloned *NORAD* downstream of a CAG promoter in a destination vector expressing BFP linked by IRES to a hygromycin resistance cassette driven by an EF1α promoter (ClonTech). 5′-truncated *NORAD* was generated by deleting bases 33–898 from *NORAD* in pDONR221 using site-directed mutagenesis (Q5 Site-Directed Mutagenesis Kit, New England Biolabs). Sequence-verified 5′-truncated *NORAD* pDONR221 was cloned into the described destination vector using LR recombination.

Sequence-verified rescue constructs were transfected into CRISPRi cells with stably integrated *NORAD* sgRNAs using FuGene HD (Promega) by following the manufacturer's instructions. We selected cells that stably integrated *NORAD* rescue constructs by selecting with hygromycin B (Sigma Aldrich) at a final concentration of 25 μg/ml. Knockdown of endogenous *NORAD* was achieved by inducing KRAB–dCas9 expression in CRISPRi cell lines stably expressing sgRNAs targeting the endogenous *NORAD* promoter using doxycycline at 0.5 μg/ml. RT–qPCR Primers specific to the 5′ end of *NORAD* (forward primer: CTCTGCTGTGGCTGCCC, reverse primer: GGGTGGGAAAGAGAGGTTCG) or a middle segment of *NORAD* (forward primer: CTCTCCACCACCAACCTGATG, reverse primer: GGAAGTGAGATAACATCAGCTCTAA) were used to verify expression of full-length or 5′-truncated *NORAD* in cells depleted of endogenous *NORAD*.

Cell-cycle analysis. Cell-cycle analysis was carried out by measuring EdU incorporation and total DNA content. CRISPRi cells with stably integrated *NORAD* or *RBMX* sgRNAs and stably integrated rescue cassettes expressing different *NORAD* constructs (full-length *NORAD*, 5′-truncated *NORAD* or empty rescue cassette) were maintained in medium containing hygromycin B (Sigma Aldrich) at a final concentration of 12.5 μg/ml. Induction of KRAB–dCas9 and constitutive expression of rescue cassettes was routinely monitored by fluorescence-activated cell sorting. Medium supplemented with 0.5 μg/ml doxycycline was added to knockdown samples for 48 h. We then trypsinized cells and plated them in 24-well cell culture plates using 100,000 cells per well and incubated them for another 24 h in the presence of doxycycline. We labelled newly replicating DNA by supplementing cell culture medium with 10 μM EdU for 1 h. Cells were washed with PBS, trypsinized and transferred to a 96-well round-bottom plate for improved handling of many samples in parallel. We used the Click-iT Plus EdU Flow Cytometry Assay Kit (Thermo Fisher Scientific) and followed the manufacturer's instructions with the following modifications. For improved multiplexing, we reduced the number of cells per assay by a factor of 10 (1×10^6 cells/ml) and scaled down washing volumes accordingly. The Click-IT reaction was performed using half the recommended reagent volumes per sample. After the last washing step, cells were resuspended in PBS containing FxCycle Far Red Stain (Thermo Fisher Scientific) as well as RNase Cocktail (Thermo Fisher Scientific) and incubated for 30 min at room temperature to stain total DNA. Fluorescence-activated cell sorting was performed on a CytoFLEX S Instrument (Beckman Coulter).

RNAi knockdown experiments. For RNAi knockdown experiments, we plated 50,000 HCT116 cells 24 h before transfection into 24-well tissue culture plates using antibiotic-free medium. We transfected 50 nM short interfering RNAs (siRNAs) into each well using Lipofectamine RNAiMAX (Thermo Fisher Scientific) by following the manufacturer's instructions. Medium was changed the day after transfections and cells were incubated with siRNAs for a total time of 72 h. Cell-cycle analysis and RT–qPCR were performed as described in the above sections (see 'Cell-cycle analysis' and 'RNA extraction and RT–qPCR' above).

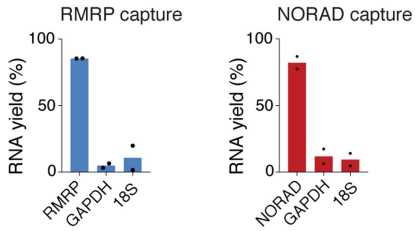
DNA combing. We induced knockdown in CRISPRi cells with stably integrated sgRNAs by supplementing cell culture medium with 0.5 μg/ml doxycycline for 72 h. Knockdown and wild-type cells were labelled with a final concentration of 100 μM CIdU for 70 min. CIdU-containing medium was removed, cells were washed twice with warm medium (which did not contain any thymidine analogues) and then incubated with a final concentration of 100 μM IdU for 70 min. IdU-containing medium was removed, cells were washed twice with warm PBS and trypsinized. We counted cells in triplicates and used 75,000 cells for each experiment. Cells were embedded in agarose plugs using the FibrePrep DNA Extraction Kit (Genomic Vision) by following the manufacturer's instructions. DNA combing, immunodetection, image acquisition and data analysis were performed at specialized service facilities. Only intact replication origins with positive DNA counterstaining were used to measure fibre length and calculate replication fork velocity.

Reporting summary. Further information on research design is available in the Nature Research Reporting Summary linked to this paper.

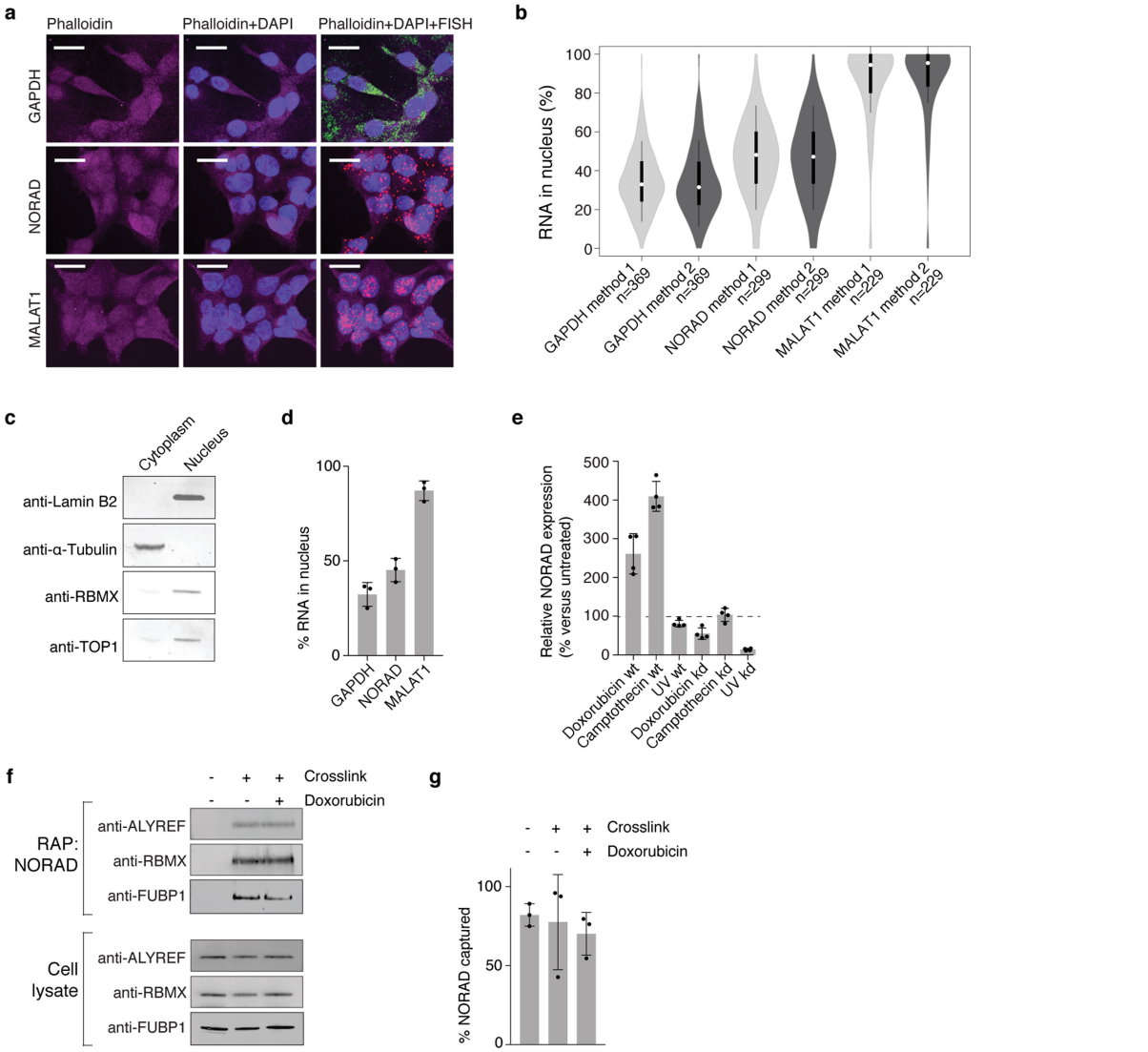
Code availability. Code for the analyses described in this paper is available from the corresponding authors upon request.

Data availability. Sequencing data for this study are available at the Gene Expression Omnibus under the accession number GSE114953. The original mass spectra may be downloaded from MassIVE (http://massive.ucsd.edu) using the identifier: MSV000082561. The data are directly accessible via ftp://massive.ucsd.edu/MSV000082561. All other data are available from the corresponding authors upon reasonable request.

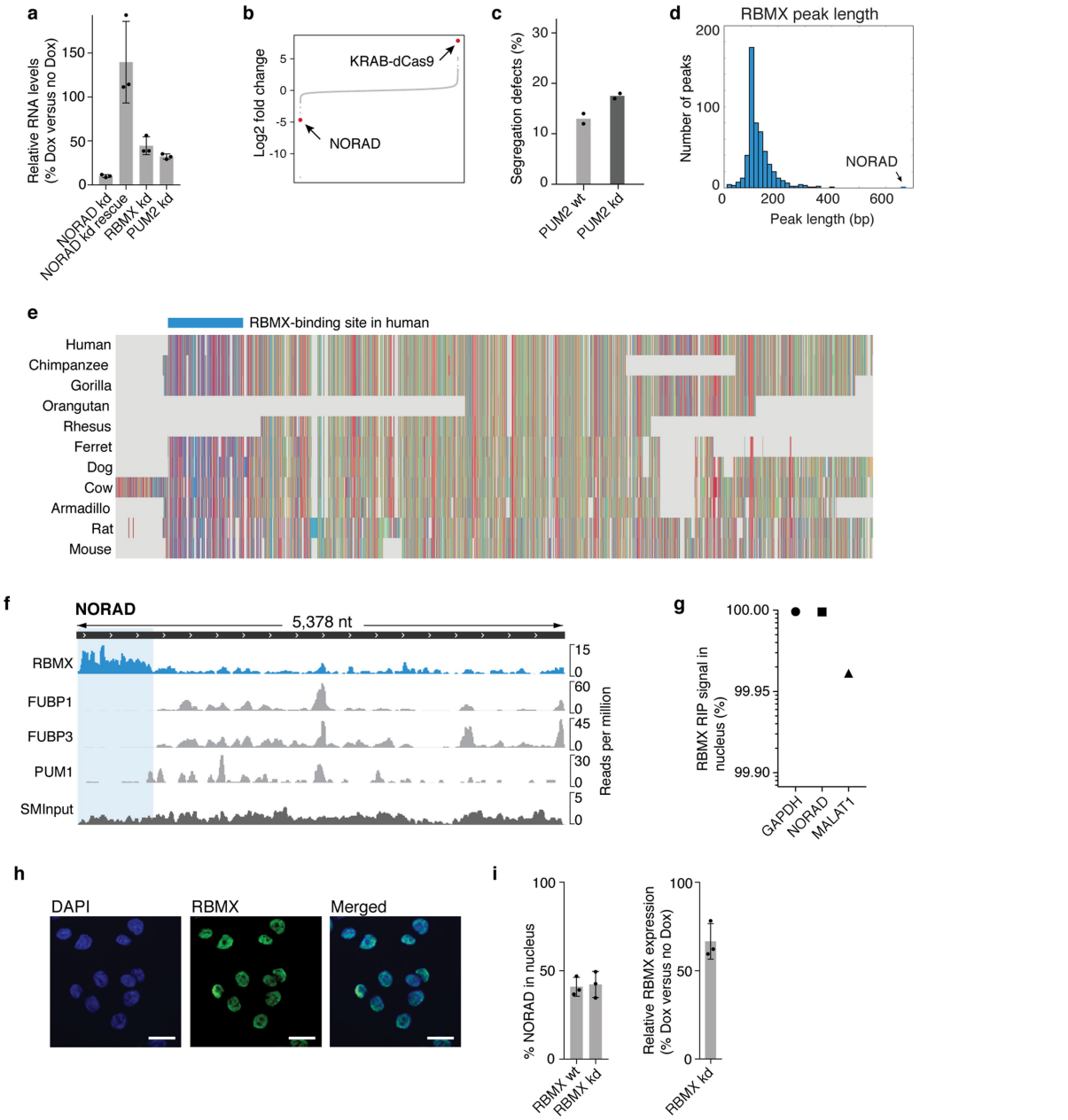
31. Fulco, C. P. et al. Systematic mapping of functional enhancer–promoter connections with CRISPR interference. *Science* **354**, 769–773 (2016).
32. Engreitz, J. M. et al. The Xist lncRNA exploits three-dimensional genome architecture to spread across the X chromosome. *Science* **341**, 1237973 (2013).
33. Rappsilber, J., Ishihama, Y. & Mann, M. Stop and go extraction tips for matrix-assisted laser desorption/ionization, nanoelectrospray, and LC/MS sample pretreatment in proteomics. *Anal. Chem.* **75**, 663–670 (2003).
34. Wiese, S., Reidegeld, K. A., Meyer, H. E. & Warscheid, B. Protein labeling by iTRAQ: a new tool for quantitative mass spectrometry in proteome research. *Proteomics* **7**, 340–350 (2007).
35. Smyth, G. K. Linear models and empirical Bayes methods for assessing differential expression in microarray experiments. *Stat. Appl. Genet. Mol. Biol.* **3**, article 3 (2004).
36. Engreitz, J. M. et al. RNA–RNA interactions enable specific targeting of noncoding RNAs to nascent pre-mRNAs and chromatin sites. *Cell* **159**, 188–199 (2014).
37. Zarnegar, B. J. et al. irCLIP platform for efficient characterization of protein-RNA interactions. *Nat. Methods* **13**, 489–492 (2016).
38. Hafner, M. et al. Transcriptome-wide identification of RNA-binding protein and microRNA target sites by PAR-CLIP. *Cell* **141**, 129–141 (2010).
39. Van Nostrand, E. L. et al. Robust transcriptome-wide discovery of RNA-binding protein binding sites with enhanced CLIP (eCLIP). *Nat. Methods* **13**, 508–514 (2016).
40. Kim, D. et al. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol.* **14**, R36 (2013).
41. Zhang, Y. et al. Model-based analysis of ChIP-seq (MACS). *Genome Biol.* **9**, R137 (2008).
42. Bray, N. L., Pimentel, H., Melsted, P. & Pachter, L. Near-optimal probabilistic RNA-seq quantification. *Nat. Biotechnol.* **34**, 525–527 (2016).
43. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).
44. Trincado, J. L. et al. SUPPA2: fast, accurate, and uncertainty-aware differential splicing analysis across multiple conditions. *Genome Biol.* **19**, 40 (2018).
45. Hu, Y. et al. DiffSplice: the genome-wide detection of differential splicing events with RNA-seq. *Nucleic Acids Res.* **41**, e39 (2013).
46. GTEx Consortium. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. *Science* **348**, 648–660 (2015).
47. Brawand, D. et al. The evolution of gene expression levels in mammalian organs. *Nature* **478**, 343–348 (2011).
48. Merkin, J., Russell, C., Chen, P. & Burge, C. B. Evolutionary dynamics of gene and isoform regulation in mammalian tissues. *Science* **338**, 1593–1599 (2012).
49. Chen, J. et al. Evolutionary analysis across mammals reveals distinct classes of long non-coding RNAs. *Genome Biol.* **17**, 19 (2016).
50. Dobin, A. et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
51. Zerbino, D. R. et al. Ensembl 2018. *Nucleic Acids Res.* **46**, D754–D761 (2017).
52. Pertea, M. et al. StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat. Biotechnol.* **33**, 290–295 (2015).
53. Katoh, K. & Standley, D. M. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* **30**, 772–780 (2013).
54. Mueller, F. et al. FISH-quant: automatic counting of transcripts in 3D FISH images. *Nat. Methods* **10**, 277–278 (2013).
55. Kamentsky, L. et al. Improved structure, function and compatibility for CellProfiler: modular high-throughput image analysis software. *Bioinformatics* **27**, 1179–1180 (2011).



Extended Data Fig. 1 | RNA antisense purification of *RMRP* and *NORAD* transcripts. RT–qPCR measurements of RNA yield in *RMRP* and *NORAD* RAP MS experiments. Columns represent the mean of two biological replicate experiments, individual data points are shown.



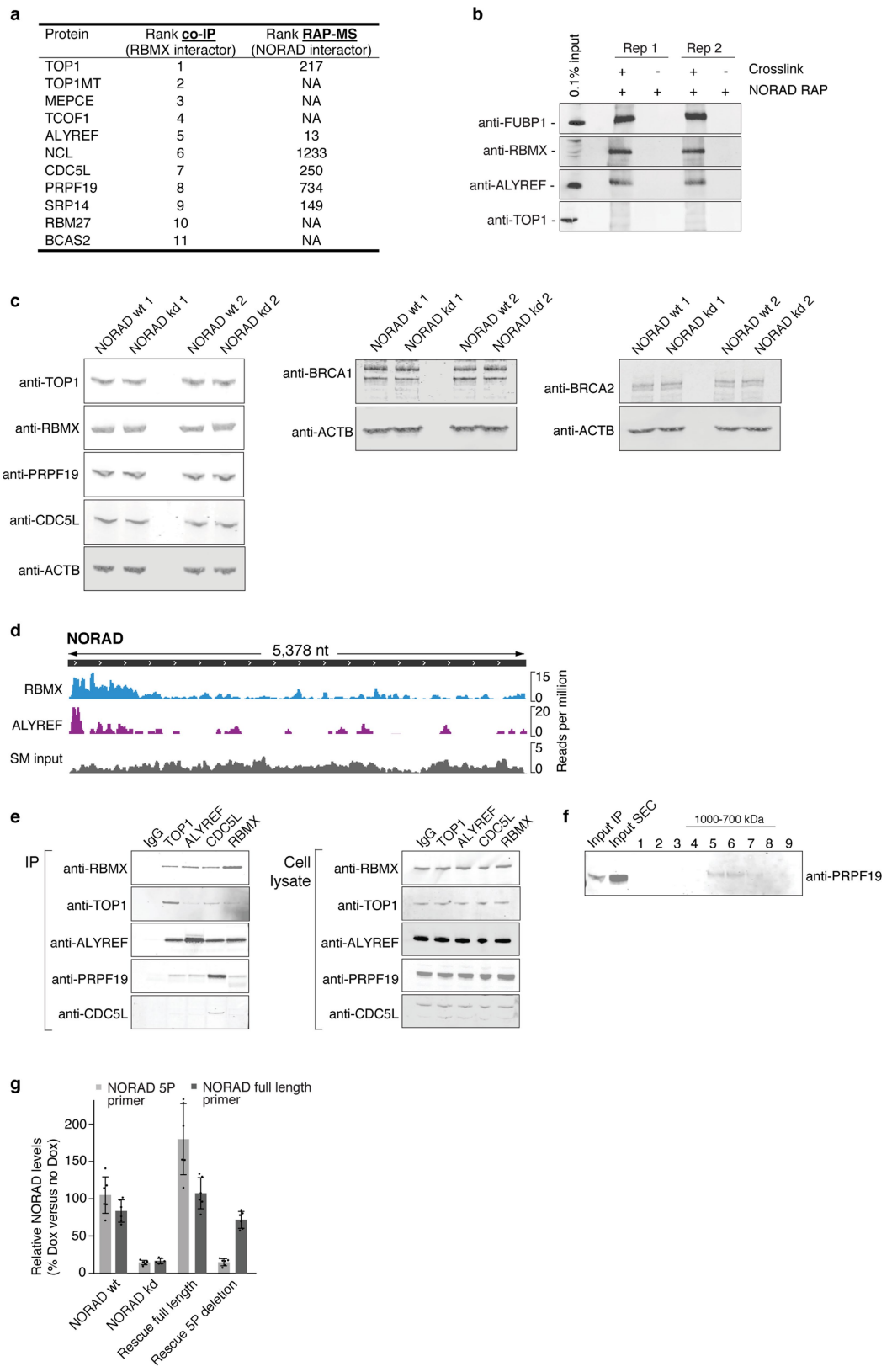
Extended Data Fig. 2 | Subcellular localization of *NORAD* and analysis of *NORAD*–protein interactions with DNA damage. a, smRNA FISH of *GAPDH*, *NORAD* and *MALAT1* in wild-type HCT116 cells. *GAPDH*, cytoplasmic reference; *MALAT1*, nuclear reference. Actin is stained with Alexa Fluor 647-conjungated phalloidin. Scale bar, 20 μm. Images are representative of one experiment; three independent experiments were performed. **b**, Quantification of smRNA FISH experiments. Circles show medians; box limits, 25th and 75th percentiles; whiskers, 1.5× interquartile range; polygons, extreme values. Method 1: phalloidin-aided cell boundary detection using the watershed method. Method 2: proximity-based cell boundary detection using the distance − N method in Cell Profiler. Sample sizes: *GAPDH* method 1, $n = 369$; *GAPDH* method 2, $n = 369$; *NORAD* method 1, $n = 299$; *NORAD* method 2, $n = 299$; *MALAT1* method 1, $n = 229$; *MALAT1* method 2, $n = 229$. **c**, Subcellular fractionation of HCT116 cells. Lamin B2 and α-tubulin serve as controls for nuclear and cytoplasmic fractions, respectively. Western blots are representative of one experiment; three independent experiments were performed. **d**, RT–qPCR measurements of relative RNA levels in nuclear and cytoplasmic extracts. Quantification relative to *GAPDH*. Percent nuclear extract is calculated relative to the total signal observed in nuclear and cytoplasmic fractions. Values are mean ± standard deviation ($n = 3$). **e**, RT–qPCR measurements of *NORAD* expression upon doxorubicin, camptothecin or ultraviolet treatment in *NORAD* wild-type or knockdown cells. Quantification relative to *GAPDH*. Values are mean ± standard deviation ($n = 4$). **f**, Western blot of *NORAD* RAP experiments with or without DNA damage. Western blots are representative of one experiment; three independent experiments were performed. **g**, RT–qPCR measurements of RNA yield in *NORAD* RAP experiments. Values are mean ± standard deviation ($n = 3$).



Extended Data Fig. 3 | Analysis of *NORAD* knockdown, *NORAD* conservation and *NORAD*–protein interactions. a, RT–qPCR measurements of *NORAD*, *RBMX* and *PUM2* CRISPRi knockdown and *NORAD* rescue experiments. Quantification relative to *GAPDH*. Values are mean ± standard deviation ($n = 3$). **b**, Differentially expressed genes in RNA-sequencing experiments from *NORAD* CRISPRi knockdown cells. **c**, Quantification of chromosome segregation errors in PUM2 wild-type or knockdown cells. One hundred anaphases were scored for each condition. Columns represent the mean of two biological replicate experiments, individual data points are shown. **d**, Histogram of RBMX-binding-site length in CLIP experiments. **e**, Multiple sequence alignment of *NORAD* transcripts, assembled de novo from RNA-sequencing data from 11 mammalian species. Only transcribed sequences are shown. Blue bar indicates RBMX-binding site in human *NORAD*. Alignment colour scheme: A, orange; C, blue; T, green; G, red. **f**, CLIP data plotted across *NORAD* RNA for RBMX, FUBP1, FUBP3 and PUM1. RBMX SM input library is shown. Representative alignments from two biological replicates are shown. **g**, RBMX RIP in nuclear and cytoplasmic fractions. The percentage of nuclear RIP signal is calculated relative to the total signal observed in nuclear and cytoplasmic fractions. **h**, Immunofluorescence imaging of RBMX in HCT116 cells. Scale bar, 20 μm. Representative images from three biological replicates are shown. **i**, Left, RT–qPCR measurements of *NORAD* RNA levels in nuclear and cytoplasmic extracts under RBMX CRISPRi wild-type or knockdown conditions. The percentage of nuclear *NORAD* is calculated relative to the total signal observed in nuclear and cytoplasmic fractions. Right, RT–qPCR measurements of *RBMX* CRISPRi knockdown. Quantification relative to *GAPDH*. Values are mean ± standard deviation ($n = 3$).

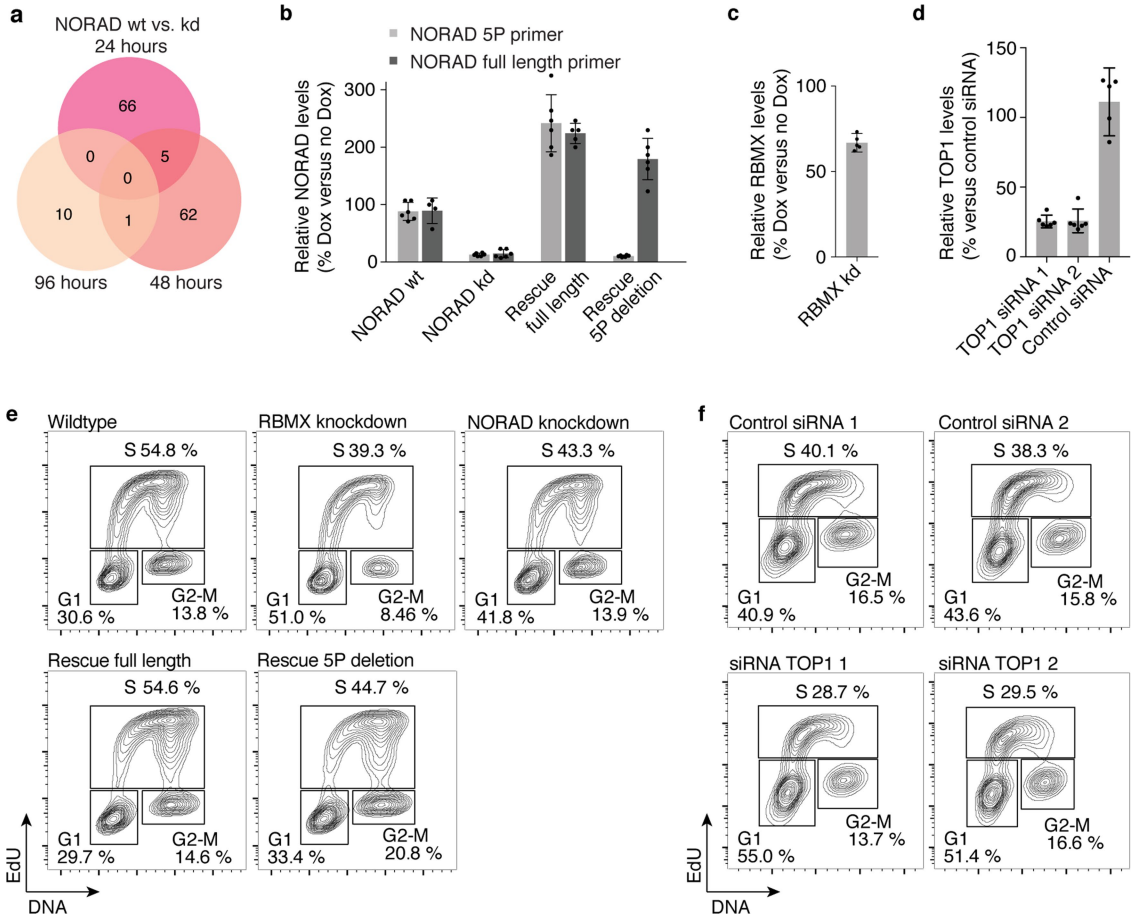
a

Protein	Rank co-IP (RBMX interactor)	Rank RAP-MS (NORAD interactor)
TOP1	1	217
TOP1MT	2	NA
MEPCE	3	NA
TCOF1	4	NA
ALYREF	5	13
NCL	6	1233
CDC5L	7	250
PRPF19	8	734
SRP14	9	149
RBM27	10	NA
BCAS2	11	NA



Extended Data Fig. 4 | See next page for caption.

Extended Data Fig. 4 | Analysis of RBMX protein–protein interactions and their dependency on *NORAD*. **a**, Ranked list of *NORAD*-dependent RBMX-interacting proteins identified by quantitative co-IP MS (Supplementary Table 4) and their respective rank in *NORAD* RAP MS experiments. **b**, Western blot of two independent *NORAD* RAP experiments with or without crosslink. Antibodies were pooled and incubated with the same membrane. Corresponding size regions were cropped for simplicity of presentation. **c**, Western blot of levels of TOP1, RBMX, PRPF19, CDC5L, BRCA1 and BRCA2 proteins in *NORAD* wild-type and knockdown cells from two independent experiments. β-actin serves as loading control. **d**, CLIP data plotted across *NORAD* RNA for RBMX and ALYREF. RBMX SM input library is shown. Representative alignments from two biological replicates are shown. **e**, Co-IP western blot for TOP1, ALYREF, CDC5L, RBMX and IgG control. Inputs are shown on the right. Western blots are representative of one experiment; three independent experiments were performed. **f**, Western blot of Flag–RBMX–V5 co-IP followed by size-exclusion chromatography. Fractions 1–9 are shown. Fractions 10–20 were not probed for PRPF19 owing to overlap with Flag antibody at this size range (Supplementary Note 4). **g**, RT–qPCR measurements of *NORAD* 5′ fragment (light grey) and full-length *NORAD* (dark grey) in rescue experiments using full-length and 5′-truncated *NORAD* rescue constructs. Measurements correspond to cells used for proximity ligation assays. Quantification relative to *GAPDH*. Values are mean ± standard deviation ($n = 6$).



Extended Data Fig. 5 | Analysis of alternative splicing and cell-cycle progression in *NORAD* depleted cells. a, Venn diagram of significant splicing changes (percentage spliced in (PSI) > 20%; FDR < 0.05) in *NORAD* wild-type and knockdown cells at 24, 48 and 96 h (Supplementary Table 6); 89,352, 88,529 and 84,340 events were analysed at 24, 48 and 96 h, respectively. Only six events were consistent between two time points and none were consistent between all three time points **b**, RT–qPCR measurements of *NORAD* 5′ fragment (light grey) and full-length *NORAD* (dark grey) in rescue experiments using full-length and 5′-truncated *NORAD* rescue constructs. Measurements correspond to cells used in cell-cycle analysis. Quantification relative to *GAPDH*. Values are mean ± standard deviation ($n = 5$ or 6). **c**, RT–qPCR measurements of *RBMX* CRISPRi knockdown. Quantification relative to *GAPDH*. Values are mean ± standard deviation ($n = 5$). **d**, RT–qPCR measurements of *TOP1* RNA interference knockdown. Quantification relative to *GAPDH*. Values are mean ± standard deviation (*TOP1* siRNA, $n = 6$; control siRNA, $n = 5$). **e**, Representative fluorescence-activated cell sorting histograms measuring EdU incorporation and DNA content in RBMX and *NORAD* CRISPRi knockdown and *NORAD* rescue cells. Percentage of cells in each cell-cycle phase is indicated. **f**, As in **e**, but for *TOP1* RNA interference knockdown cells.

LETTER

https://doi.org/10.1038/s41586-018-0454-y

Structures of filaments from Pick's disease reveal a novel tau protein fold

Benjamin Falcon[1], Wenjuan Zhang[1], Alexey G. Murzin[1], Garib Murshudov[1], Holly J. Garringer[2], Ruben Vidal[2], R. Anthony Crowther[1], Bernardino Ghetti[2], Sjors H. W. Scheres[1,3]* & Michel Goedert[1,3]*

The ordered assembly of tau protein into abnormal filamentous inclusions underlies many human neurodegenerative diseases[1]. Tau assemblies seem to spread through specific neural networks in each disease[2], with short filaments having the greatest seeding activity[3]. The abundance of tau inclusions strongly correlates with disease symptoms[4]. Six tau isoforms are expressed in the normal adult human brain—three isoforms with four microtubule-binding repeats each (4R tau) and three isoforms that lack the second repeat (3R tau)[1]. In various diseases, tau filaments can be composed of either 3R or 4R tau, or of both. Tau filaments have distinct cellular and neuroanatomical distributions[5], with morphological and biochemical differences suggesting that they may be able to adopt disease-specific molecular conformations[6,7]. Such conformers may give rise to different neuropathological phenotypes[8,9], reminiscent of prion strains[10]. However, the underlying structures are not known. Using electron cryo-microscopy, we recently reported the structures of tau filaments from patients with Alzheimer's disease, which contain both 3R and 4R tau[11]. Here we determine the structures of tau filaments from patients with Pick's disease, a neurodegenerative disorder characterized by frontotemporal dementia. The filaments consist of residues Lys254–Phe378 of 3R tau, which are folded differently from the tau filaments in Alzheimer's disease, establishing the existence of conformers of assembled tau. The observed tau fold in the filaments of patients with Pick's disease explains the selective incorporation of 3R tau in Pick bodies, and the differences in phosphorylation relative to the tau filaments of Alzheimer's disease. Our findings show how tau can adopt distinct folds in the human brain in different diseases, an essential step for understanding the formation and propagation of molecular conformers.

We used electron cryo-microscopy (cryo-EM) to image tau filaments extracted from the frontotemporal cortex of a patient who had a 7-year history of behavioural-variant frontotemporal dementia (patient 4). Neuropathological examination revealed severe frontotemporal lobar degeneration, with abundant Pick bodies composed of 3R tau filaments, without phosphorylation of Ser262 (Fig. 1a–d, Extended Data Fig. 1, Extended Data Table 1), consistent with a diagnosis of Pick's disease[12–17]. As in Alzheimer's disease[18], a fuzzy coat composed of the disordered N- and C-terminal regions of tau surrounded the filament cores and was removed by pronase treatment (Fig. 1e and Extended Data Fig. 1). Narrow (93%) and wide (7%) filaments could be distinguished (Fig. 1e). The narrow filaments have previously been described as straight[19–21], but they do have a helical twist with a crossover distance of approximately 1,000 Å and a projected width that varies from approximately 50 to 150 Å. The wide filaments have a similar crossover distance, but their width varies from approximately 50 to 300 Å. We named them narrow and wide Pick filaments (NPFs and WPFs, respectively). Their morphologies and relative abundance match those reported in cortical biopsies from patients with Pick's disease[21].

Using helical reconstruction in RELION[22], we determined a 3.2 Å resolution map of the ordered core of NPFs, in which side-chain densities were well resolved and β-strands were clearly separated along the helical axis (Fig. 1f and Extended Data Fig. 2). We also determined a map of WPFs, which was limited to 8 Å resolution because of the small number of filaments. This was sufficient to show separated β-sheets within the structure, but not the 4.7 Å spacing of individual β-strands along the helical axis (Fig. 1g and Extended Data Fig. 3). NPFs are composed of a single protofilament with an elongated structure that is markedly different from the C-shaped protofilament of paired helical filaments (PHFs) and straight filaments found in Alzheimer's disease[11,23]. WPFs are formed by the association of two NPF protofilaments at their distal tips. In support, we observed WPFs in which one protofilament had been lost in some parts (Extended Data Fig. 3). Our results reveal that the tau filaments of Pick's disease adopt a single fold that is different from that of the tau filaments found in Alzheimer's disease.

The high-resolution NPF map allowed us to build an atomic model of the Pick tau filament fold (hereafter termed the Pick fold), which

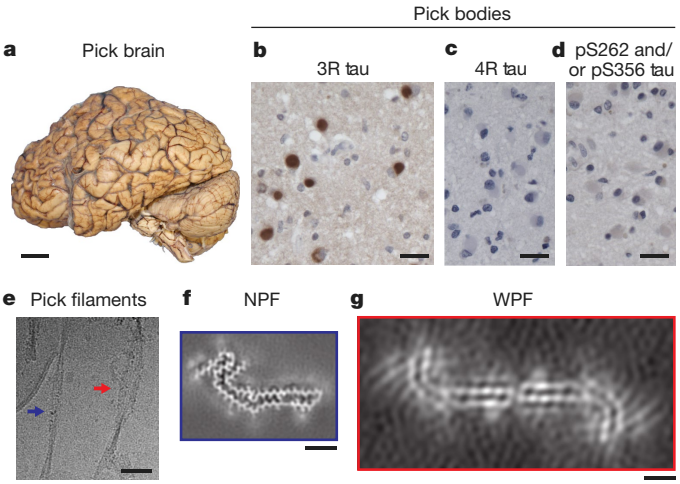


Fig. 1 | Filamentous tau pathology of Pick's disease. a, The brain used for cryo-EM (patient 4) showed atrophy of the anterior frontal and temporal lobes of the cerebral cortex. Scale bar, 5 cm. **b**–**d**, Staining of Pick bodies in the frontotemporal cortex of patient 4 using antibody RD3 (3R tau; brown) (**b**), but not by antibodies anti-4R (4R tau) (**c**) or 12E8 (pS262 tau and/or pS356 tau) (**d**). Nuclei were counterstained blue. Scale bars, 20 μm. **e**, Cryo-electron micrograph of tau filaments extracted from grey matter of the frontotemporal cortex of patient 4, in which narrow (NPFs; blue arrow) and wide (WPFs; red arrow) Pick filaments could be distinguished. Scale bar, 500 Å. **f**, Unsharpened cryo-EM density of NPF from patient 4. Scale bar, 25 Å. **g**, Unsharpened cryo-EM density of WPF from patient 4. Scale bar, 25 Å.

[1]MRC Laboratory of Molecular Biology, Cambridge, UK. [2]Department of Pathology and Laboratory Medicine, Indiana University School of Medicine, Indianapolis, IN, USA. [3]These authors jointly supervised this work: Sjors H. W. Scheres, Michel Goedert. *e-mail: scheres@mrc-lmb.cam.ac.uk; mg@mrc-lmb.cam.ac.uk

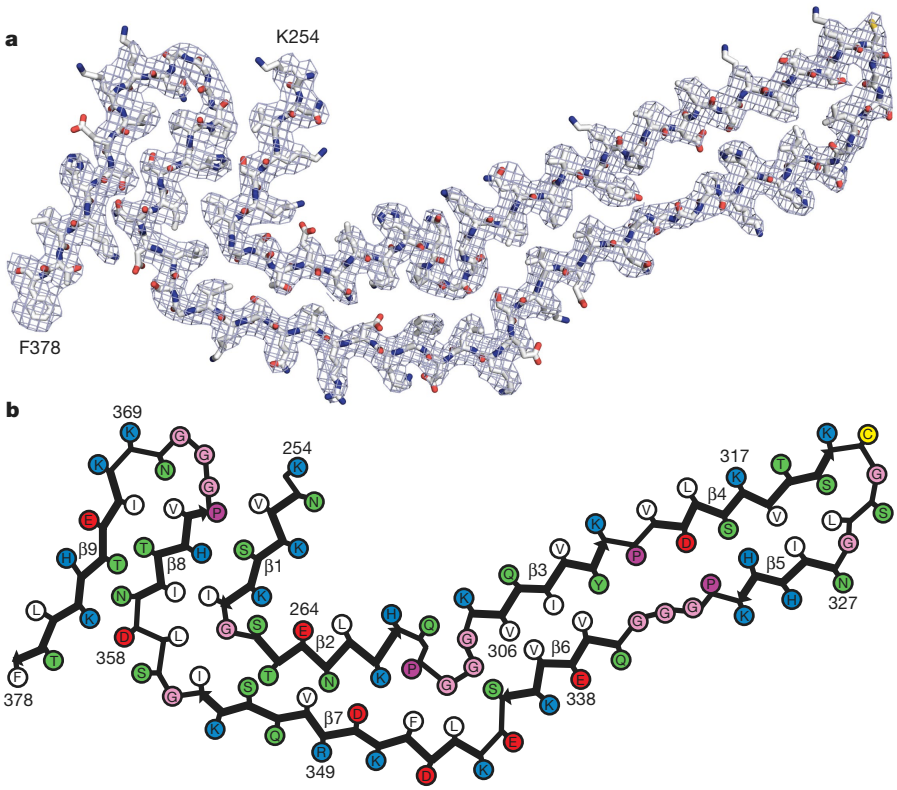


Fig. 2 | The Pick tau filament fold. **a**, Sharpened, high-resolution cryo-EM map of the NPF with the atomic model of the Pick fold overlaid. **b**, Schematic view of the Pick fold. Amino acid numbering corresponds to the 441-amino-acid human tau isoform, so residues 275–305 of R2 are not present.

consists of residues Lys254–Phe378 of 3R tau (in the numbering of the 441-amino-acid human tau isoform) (Fig. 2). There are nine β-strands (β1–β9) arranged into four cross-β packing stacks and connected by turns and arcs. R1 provides two strands, β1 and β2, and R3 and R4 provide three strands each, β3–β5 and β6–β8, respectively. These strands pack together in a hairpin-like fashion: β1 against β8, β2 against β7, β3 against β6, and β4 against β5. The final strand, β9, is formed from nine amino acids after R4 and packs against the opposite side of β8. Only the interface between β3 and β6 is entirely hydrophobic; the other cross-β packing interfaces are composed of both non-polar and polar side chains.

The interstrand connections and their interactions maintain the strand pairings and compensate for differences in strand lengths and orientations. A sharp right-angle turn at Gly261, between β1 and β2, faces a smooth four-residue arc formed by 355-Gly-Ser-Leu-Asp-358, between β7 and β8, turning the protein chain in the same direction. The 270-Pro-Gly-Gly-Gly-273 motif between β2 and β3 gives rise to an omega-shaped turn that compacts the protein chain locally, but maintains its direction at either end. On the opposite side, a β-arc formed of Glu342 and Lys343, between β6 and β7, creates space for this turn. By contrast, the homologous 332-Pro-Gly-Gly-Gly-335 motif connecting β5 and β6 forms an extended β-spiral conformation, compensating for the shorter lengths of these strands compared to the opposing β3 and β4, which are connected by Pro312. Solvent-mediated interactions may occur within the large cavity between this motif and the side chains at the junction of β3 and β4. The third homologous 364-Pro-Gly-Gly-Gly-367 motif contributes to a 180° turn that allows β9 to pack against the other side of β8. Variations in the height of the chain along the helical axis also help to maintain an ordered hydrogen-bonding pattern of the β-stranded regions (Fig. 3c).

The solvent-exposed side chains of Cys322 and Ser324, together with the intervening Gly323, form a flat surface at the hairpin turn between β4 and β5. This provides the interface for the formation of WPFs by the abutting of protofilaments (Extended Data Fig. 3). The distances between protofilaments at this interface would enable van der Waals interactions, but not the formation of disulfide bonds. In support, WPFs were stable under reducing conditions (Extended Data Fig. 3). At the resolution of the WPF map, it was not possible to determine whether protofilament β-strands are aligned or staggered by one half of the β-strand distance, as is the case for PHFs in Alzheimer's disease[11]. We conclude that WPFs are formed by two separate protofilaments making tight contacts at their distal tips through van der Waals interactions.

Three regions of less well-resolved density bordering the solvent-exposed faces of β4, β5 and β9 are apparent in the unsharpened maps of both NPFs and WPFs (Fig. 1f, g). Their low resolution suggests that they represent less ordered, heterogeneous and/or transiently occupied structures. The density bordering β4 is similarly located, but more extended and less-well resolved, than that found to interact with the side chains of Lys317, Thr319 and Lys321 in Alzheimer's disease PHFs and straight filaments[11], and hypothesized to be the N-terminal 7-Glu-Phe-Glu-9, part of the discontinuous MC1 epitope[24]. NPFs and WPFs were labelled by MC1 (Extended Data Fig. 1).

It was not previously known why only 3R tau, which lacks the second microtubule-binding repeat, is present in Pick body filaments. Our results suggest that despite sequence homology, the structure formed by Lys254–Lys274 of the first tau repeat is inaccessible to the corresponding residues from the second repeat of 4R tau (Ser285–Ser305). The site occupied by Thr263, which is in close proximity to the backbone of the opposite strand β7, cannot accommodate the bulkier side chain of Lys294 from 4R tau and does not provide a favourable environment for its charged ε-amino group (Extended Data Fig. 4). Also, the site preceding the omega-like structure formed by 270-Pro-Gly-Gly-Gly-273 cannot accommodate a C_β branched residue, such as Val300 from 4R tau instead of Gln269 from 3R tau. In addition, the smaller Cys291 residue from 4R tau would form weaker interactions with Leu357 and Ile360 than those formed by Ile260 of 3R tau. In support, tau filaments extracted from the brain of the patient with Pick's disease used for cryo-EM seeded the aggregation of recombinant full-length 3R, but not 4R, tau (Extended Data Fig. 5). Similar experiments have shown that Alzheimer's disease PHFs and straight filaments, the core sequences of which are shared by 3R and 4R tau, can seed both types of isoform[25]. Such templated misfolding explains the selective incorporation of 3R tau in Pick body filaments. Pick's disease extracts

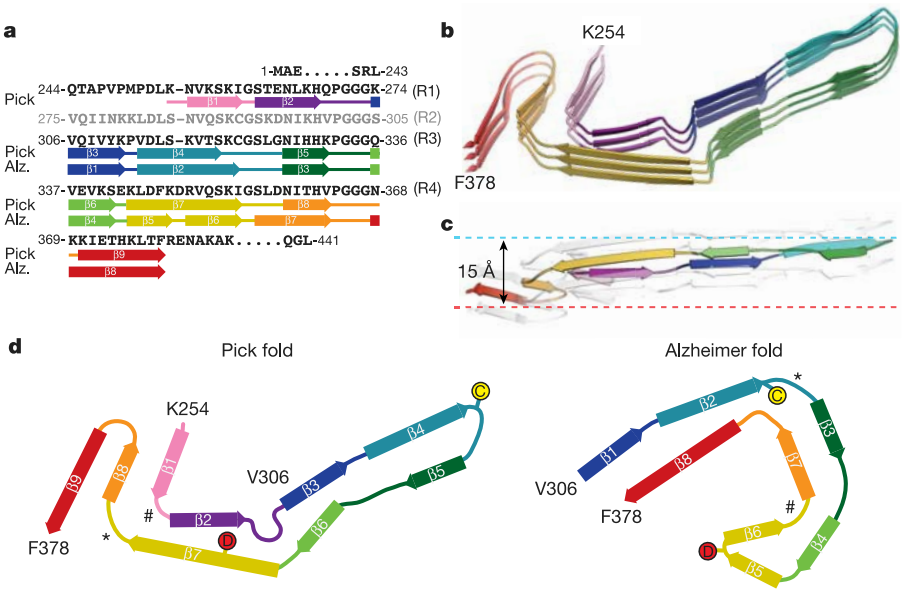


Fig. 3 | Comparison of the Pick and Alzheimer tau filament folds. **a**, Sequence alignment of the microtubule-binding repeats (R1–R4) with the observed nine β-strand regions (arrows) in the Pick fold and eight β-strand regions (arrows) in the Alzheimer fold, coloured from violet to red. **b**, Rendered view of the secondary structure elements in the Pick fold, depicted as three successive rungs. **c**, As in **b**, but in a view perpendicular to the helical axis, revealing the changes in height within a single molecule. **d**, Schematic of the secondary structure elements in the Pick and Alzheimer folds, depicted as a single rung. The positions of Cys322 (yellow 'C') and Asp348 (red 'D') in the two folds are highlighted. The asterisk and hash symbols mark conserved turns of homologous regions in the Pick and Alzheimer folds.

have been reported to seed the aggregation of a 4R tau fragment comprising the repeats (residues 244–372) with mutations Phe301Leu and Val337Met[26]. However, this tau fragment cannot form the Pick fold, which is unable to accommodate R2 and requires residues 373–378. A small amount of aggregated four-repeat tau may have accounted for the seeding activity, as suggested in a separate study[8]. Loss of von Economo neurons in anterior cingulate and frontoinsular cortices has been reported to be an early event in Pick's disease[27,28]. It remains to be established how 3R tau seeds can form in cells that also express 4R tau. Alternatively, nerve cell populations may be distinguished by the tau isoforms that they express[29].

The cryo-EM structures presented here were derived from filaments that were extracted from the brain of a single patient with Pick's disease (patient 4). To test the generality of the Pick fold, we investigated the binding of repeat-specific antibodies to tau filaments extracted from the frontotemporal cortex of eight additional patients with sporadic Pick's disease (Extended Data Table 1). By western blotting, all samples ran as two tau bands of 60 and 64 kDa, which were detected by anti-R1, anti-R3 and anti-R4 antibodies, but not by an anti-R2 antibody, showing the presence of only 3R tau (Extended Data Fig. 6). Immunogold negative-stain electron microscopy showed that most filaments were NPFs, with a minority of WPFs, and were not decorated by the repeat-specific antibodies (Extended Data Fig. 7). This shows that the R1, R3 and R4 epitopes are inaccessible to the antibodies used, indicating that they form part of the ordered filament core. Alzheimer's disease PHFs and straight filaments are decorated by anti-R2, but not by anti-R3 or anti-R4 antibodies, because their core is made of R3, R4 and the 10 amino acids following R4[11,30]. These results are in good agreement with experiments using limited proteolysis and mass spectrometry[7]. We conclude that the ordered core of tau filaments from Pick's disease comprises the C-terminal part of R1, all of R3 and R4, as well as 10 amino acids after R4.

Unlike PHFs and straight filaments in Alzheimer's disease, Pick body filaments are not phosphorylated at Ser262 and/or Ser356[14,16] (Extended Data Fig. 1). The reasons for this differential phosphorylation are unknown. Our structure reveals that the tight turn at Gly261 prevents phosphorylation of Ser262 in the ordered core of Pick's disease filaments, whereas the phosphorylated Ser262 is outside the ordered core of Alzheimer's disease filaments[11]. This explains the differential phosphorylation and raises the question of whether phosphorylation at Ser262 may protect against Pick's disease.

In the Pick and Alzheimer tau filament folds, most β-structure residues between Val306 and Ile354 align locally, as do the connecting segments of Pro312, 332-Pro-Gly-Gly-Gly-335 and 342-Glu-Lys-343 (Fig. 3). Almost all amino acid side chains from this region have the same interior or solvent-exposed orientations in both folds. Exceptions are Cys322 and Asp348, which cause reversed chain directions in one or other fold (Fig. 3d). The side chain of Cys322 is interior in the Alzheimer tau filament fold, whereas it is solvent-exposed in the Pick fold. This enables the hairpin-like turn and the cross-β packing of β4 against β5. The side chain of Asp348 is interior in the Pick tau filament fold, thereby maintaining the β-structure from Lys343 to Ile354 (β7), whereas it is solvent-exposed in the Alzheimer fold, enabling the tight turn between β5 and β6, which, together with β4, gives rise to a triangular β-helix conformation[11]. Such β-helices, previously thought to be important for propagation[31], are absent from the Pick tau filament fold. The β-strands in Gly355–Phe378 align well in both folds, but have different cross-β packing arrangements. The solvent-exposed side chains of β7 and β8 in the Alzheimer fold are interior in the equivalent strands of the Pick fold (β8 and β9), because of different conformations of the two turn regions in R4, 355-Gly-Ser-Leu-Asp-358 and 364-Pro-Gly-Gly-Gly-367. The 355-Gly-Ser-Leu-Asp-358 motif makes a sharp right-angle turn at Gly355 in the Alzheimer tau filament fold, but a wide turn in the Pick fold. The same sharp turn is found at the homologous site in R1 in the Pick tau filament fold, whereas the same wide turn occurs at the homologous site in R3 in the Alzheimer fold (Fig. 3). This suggests that these semi-conserved turn structures may also be found in tau filament folds in other diseases. By contrast, the 364-Pro-Gly-Gly-Gly-367 motif adopts a new conformation in the Pick fold, which reverses the chain direction and is different from both the right-angle turn that this motif forms in the Alzheimer fold and the conformations of the homologous Pro-Gly-Gly-Gly motifs from the other repeats in both tau filament folds. The Pick and Alzheimer folds share similar secondary structure patterns, but different turn conformations result in distinct cross-β packing.

These findings show that the ordered cores of tau filaments from Pick's disease adopt a single, novel fold of 3R tau, which is distinct from

the tau filament fold found in Alzheimer's disease. This suggests that additional folds may be found in tauopathies with 4R tau filaments, such as progressive supranuclear palsy. Our results also suggest that single, disease-specific folds may exist in tauopathies with the same tau filament isoform composition, such as progressive supranuclear palsy and corticobasal degeneration, since identical tau sequences can adopt more than one fold. Conserved secondary structure motifs and markedly different conformations at turn residues in the Alzheimer and Pick tau filament folds may form the basis for structural diversity in tau protein folds from other neurodegenerative diseases. The structures of different tau filament folds will provide the basis for determining the binding sites of tau ligands and give a rationale for their interactions with inclusions in different tauopathies[32].

The identification of disease-specific folds in the ordered cores of tau filaments establishes the existence of molecular conformers. This is central to the hypothesis that conformers of filamentous tau give rise to the clinical phenotypes that define distinct tauopathies, akin to prion strains. By revealing the structural basis for molecular conformers in specific diseases, our results pave the way to a better understanding of a wide range of diseases related to abnormal protein aggregation.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, statements of data availability and associated accession codes are available at https://doi.org/10.1038/s41586-018-0454-y.

Received: 13 April 2018; Accepted: 18 July 2018;
Published online 29 August 2018.

1. Goedert, M., Eisenberg, D. S. & Crowther, R. A. Propagation of tau aggregates and neurodegeneration. *Annu. Rev. Neurosci.* **40**, 189–210 (2017).
2. Braak, H. & Del Tredici, K. Potential pathways of abnormal tau and α-synuclein dissemination in sporadic Alzheimer's and Parkinson's diseases. *Cold Spring Harb. Perspect. Biol.* **8**, a023630 (2016).
3. Jackson, S. J. et al. Short fibrils constitute the major species of seed-competent tau in the brains of mice transgenic for human P301S Tau. *J. Neurosci.* **36**, 762–772 (2016).
4. Wilcock, G. K. & Esiri, M. M. Plaques, tangles and dementia. A quantitative study. *J. Neurol. Sci.* **56**, 343–356 (1982).
5. Ghetti, B. et al. Invited review: Frontotemporal dementia caused by microtubule-associated protein tau gene (MAPT) mutations: a chameleon for neuropathology and neuroimaging. *Neuropathol. Appl. Neurobiol.* **41**, 24–46 (2015).
6. Crowther, R. A. & Goedert, M. Abnormal tau-containing filaments in neurodegenerative diseases. *J. Struct. Biol.* **130**, 271–279 (2000).
7. Taniguchi-Watanabe, S. et al. Biochemical classification of tauopathies by immunoblot, protein sequence and mass spectrometric analyses of sarkosyl-insoluble and trypsin-resistant tau. *Acta Neuropathol.* **131**, 267–280 (2016).
8. Clavaguera, F. et al. Brain homogenates from human tauopathies induce tau inclusions in mouse brain. *Proc. Natl Acad. Sci. USA* **110**, 9535–9540 (2013).
9. Narasimhan, S. et al. Pathological tau strains from human brains recapitulate the diversity of tauopathies in nontransgenic mouse brain. *J. Neurosci.* **37**, 11406–11423 (2017).
10. Collinge, J. Mammalian prions and their wider relevance in neurodegenerative diseases. *Nature* **539**, 217–226 (2016).
11. Fitzpatrick, A. W. P. et al. Cryo-EM structures of tau filaments from Alzheimer's disease. *Nature* **547**, 185–190 (2017).
12. Rasool, C. G. & Selkoe, D. J. Sharing of specific antigens by degenerating neurons in Pick's disease and Alzheimer's disease. *N. Engl. J. Med.* **312**, 700–705 (1985).
13. Pollock, N. J., Mirra, S. S., Binder, L. I., Hansen, L. A. & Wood, J. G. Filamentous aggregates in Pick's disease, progressive supranuclear palsy, and Alzheimer's disease share antigenic determinants with microtubule-associated protein, tau. *Lancet* **328**, 1211 (1986).
14. Probst, A., Tolnay, M., Langui, D., Goedert, M. & Spillantini, M. G. Pick's disease: hyperphosphorylated tau protein segregates to the somatoaxonal compartment. *Acta Neuropathol.* **92**, 588–596 (1996).
15. Delacourte, A. et al. Specific pathological tau protein variants characterize Pick's disease. *J. Neuropathol. Exp. Neurol.* **55**, 159–168 (1996).
16. Delacourte, A., Sergeant, N., Wattez, A., Gauvreau, D. & Robitaille, Y. Vulnerable neuronal subsets in Alzheimer's and Pick's disease are distinguished by their tau isoform distribution and phosphorylation. *Ann. Neurol.* **43**, 193–204 (1998).
17. Munoz, D. G., Morris, H. R. & Rossor, M. in *Neurodegeneration: the Molecular Pathology of Dementia and Movement Disorders* (eds Dickson, D. W. & Weller, R. O.) 156–164 (Wiley-Blackwell, New Jersey, 2011).
18. Wischik, C. M. et al. Structural characterization of the core of the paired helical filament of Alzheimer disease. *Proc. Natl Acad. Sci. USA* **85**, 4884–4888 (1988).
19. Rewcastle, N. B. & Ball, M. J. Electron microscopic structure of the "inclusion bodies" in Pick's disease. *Neurology* **18**, 1205–1213 (1968).
20. Schochet, S. S., Jr., Lampert, P. W. & Lindenberg, R. Fine structure of the Pick and Hirano bodies in a case of Pick's disease. *Acta Neuropathol.* **11**, 330–337 (1968).
21. Wiśniewski, H. M., Coblentz, J. M. & Terry, R. D. Pick's disease. A clinical and ultrastructural study. *Arch. Neurol.* **26**, 97–108 (1972).
22. He, S. & Scheres, S. H. W. Helical reconstruction in RELION. *J. Struct. Biol.* **198**, 163–176 (2017).
23. Crowther, R. A. Straight and paired helical filaments in Alzheimer disease have a common structural unit. *Proc. Natl Acad. Sci. USA* **88**, 2288–2292 (1991).
24. Jicha, G. A., Bowser, R., Kazam, I. G. & Davies, P. Alz-50 and MC-1, a new monoclonal antibody raised to paired helical filaments, recognize conformational epitopes on recombinant tau. *J. Neurosci. Res.* **48**, 128–132 (1997).
25. Guo, J. L. et al. Unique pathological tau conformers from Alzheimer's brains transmit tau pathology in nontransgenic mice. *J. Exp. Med.* **213**, 2635–2654 (2016).
26. Sanders, D. W. et al. Distinct tau prion strains propagate in cells and mice and define different tauopathies. *Neuron* **82**, 1271–1288 (2014).
27. Seeley, W. W. et al. Early frontotemporal dementia targets neurons unique to apes and humans. *Ann. Neurol.* **60**, 660–667 (2006).
28. Miki, Y. et al. An autopsy case of incipient Pick's disease: immunohistochemical profile of early-stage Pick body formation. *Neuropathology* **34**, 386–391 (2014).
29. Goedert, M., Spillantini, M. G., Potier, M. C., Ulrich, J. & Crowther, R. A. Cloning and sequencing of the cDNA encoding an isoform of microtubule-associated protein tau containing four tandem repeats: differential expression of tau protein mRNAs in human brain. *EMBO J.* **8**, 393–399 (1989).
30. Goedert, M., Spillantini, M. G., Jakes, R., Rutherford, D. & Crowther, R. A. Multiple isoforms of human microtubule-associated protein tau: sequences and localization in neurofibrillary tangles of Alzheimer's disease. *Neuron* **3**, 519–526 (1989).
31. Wasmer, C. et al. Amyloid fibrils of the HET-s(218-289) prion form a β solenoid with a triangular hydrophobic core. *Science* **319**, 1523–1526 (2008).
32. Goedert, M., Yamaguchi, Y., Mishra, S. K., Higuchi, M. & Sahara, N. Tau filaments and the development of positron emission tomography tracers. *Front. Neurol.* **9**, 70 (2018).

Acknowledgements We thank the patients' families for donating brain tissue; M. R. Farlow for clinical evaluation; F. Epperson, R. M. Richardson and U. Kuederli for human brain collection and analysis; P. Davies, P. Seubert and M. Hasegawa for antibodies MC1, 12E8 and TauC4, respectively; S. Chen, C. Savva and G. Cannone for support with electron microscopy; T. Darling and J. Grimmett for help with computing; W. W. Seeley and M. G. Spillantini for discussions. M.G. is an Honorary Professor in the Department of Clinical Neurosciences of the University of Cambridge. This work was supported by the UK Medical Research Council (MC_UP_A025_1012 to G.M., MC_UP_A025_1013 to S.H.W.S. and MC_U105184291 to M.G.), the European Union (Joint Programme-Neurodegeneration Research REfrAME to M.G. and B.F. and the Innovative Medicines Initiative 2 IMPRiND, project number 115881, to M.G.), the US National Institutes of Health (grant P30-AG010133 to B.G.), the Department of Pathology and Laboratory Medicine, Indiana University School of Medicine (to B.G.) and an Alzheimer's Association Zenith Award (to R.V.).

Reviewer information *Nature* thanks E. H. Egelman, B. T. Hyman and the other anonymous reviewer(s) for their contribution to the peer review of this work.

Author contributions B.G. performed neuropathology; H.J.G. and R.V. carried out genetic analysis; B.F. extracted tau filaments; B.F. and W.Z. conducted immunolabelling; B.F. and W.Z. purified recombinant tau proteins; B.F. carried out seeded aggregation; B.F. and W.Z. performed cryo-EM; B.F., W.Z. and S.H.W.S. analysed the cryo-EM data; B.F., W.Z., G.M. and A.G.M. built the atomic model; R.A.C. contributed to the inception of the study; M.G. and S.H.W.S. supervised the project; all authors contributed to writing the manuscript.

Competing interests : The authors declare no competing interests.

Additional information
Extended data is available for this paper at https://doi.org/10.1038/s41586-018-0454-y.
Supplementary information is available for this paper at https://doi.org/10.1038/s41586-018-0454-y.
Reprints and permissions information is available at http://www.nature.com/reprints.
Correspondence and requests for materials should be addressed to S.H.W.S. or M.G.
Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

METHODS

No statistical methods were used to predetermine sample size. The experiments were not randomized, and investigators were not blinded to allocation during experiments and outcome assessment.

Extraction of tau filaments. Sarkosyl-insoluble material was extracted from the grey matter of frontal and temporal cortex of the patients' brains, as described[33]. Approximately 6 g tissue was used for cryo-EM and 0.6 g tissue was used for immunolabelling. The pelleted sarkosyl-insoluble material was resuspended in 50 mM Tris-HCl pH 7.4 containing 150 mM NaCl and 0.02% amphipol A8-35 at 250 μl per g tissue, followed by centrifugation at 3,000*g* for 30 min at 4 °C. The pellets, containing large contaminants, were discarded. The supernatants were centrifuged at 100,000*g* for 30 min at 4 °C. The resulting pellets were resuspended in buffer at 15 μl per gram tissue for cryo-EM and 150 μl per g tissue for immunolabelling. Pronase treatment was carried out as described for negative-stain EM[33] and cryo-EM[11]. Tau filaments were incubated with 0.4 mg ml^{-1} pronase (Sigma) for 5 min at 21 °C before being deposited on EM grids and plunge-frozen.

Recombinant tau. Tau constructs lacking the BR136, anti-4R, BR135 and TauC4 peptide sequences were cloned from pRK172 encoding wild-type 0N4R or 2N4R tau using the QuikChange Lightning site-directed mutagenesis kit (Agilent), according to the manufacturer's instructions. All recombinant tau proteins were expressed and purified as described[34].

Immunolabelling and histology. Western blotting and immunogold negative-stain EM were carried out as described previously[33]. For western blotting, samples were resolved on 4–20% or 10% Tris-glycine gels (Novex), and the primary antibodies were diluted in PBS plus 0.1% Tween 20 and 1% bovine serum albumin (BSA). BR136 is a polyclonal antibody that was raised against a synthetic peptide corresponding to residues 244–257 of tau. The peptide (200 μg), coupled to keyhole limpet haemocyanin using glutaraldehyde, was mixed 1:1 with Freund's complete adjuvant and used to immunize white Dutch rabbits. Booster injections using 200 μg of conjugated peptide mixed 1:1 with Freund's incomplete adjuvant were given every 2 weeks for 10 weeks after the primary immunization. Antibodies were obtained 7 days after the final booster injection and affinity purified. Extended Data Fig. 6 shows that BR136 is specific for the C-terminal region of residues 244–257. Neurohistology and immunohistochemistry were carried out as described previously[35]. Brain sections were 8-μm thick and were counterstained with haematoxylin. Detailed antibody information is provided in Extended Data Table 2.

Whole-exome sequencing. Whole-exome sequencing was carried out at the Center for Medical Genomics of Indiana University School of Medicine using genomic DNA from the nine individuals with neuropathologically confirmed diagnoses of Pick's disease, the tau filaments of which were used in Extended Data Figs. 6 and 7. Target enrichment was performed using the SureSelectXT human all exon library (V6, 58Mb, Agilent) and high-throughput sequencing using a HiSeq4000 (2 × 75-bp paired-end configuration, Illumina). Bioinformatics analyses were performed as described[36]. Findings on *MAPT*, *PSEN1* and *APOE* are presented in Extended Data Table 1.

Seeded aggregation. Seeded aggregation was carried out as described[37], but with full-length wild-type tau protein and without the aggregation inducer heparin. Recombinant 0N3R and 0N4R tau were purified as described[34]. Extracted tau filaments (15 μl per g tissue) were diluted 1:10 in 10 mM HEPES pH7.4, 200 mM NaCl; 2 μl was added to 98 μl of 20 μM 0N3R or 0N4R recombinant tau in the same buffer with 10 μM thioflavin T in a black, clear bottom 96-well plate (Perkin Elmer). The plate was sealed and incubated at 37 °C in a plate reader (BMG Labtech FLUOstar Omega), with cycles of shaking for 60 s (500 r.p.m., orbital) followed by no shaking for 60 s. Filament formation was monitored by measuring thioflavin T fluorescence every 45 min using 450 ± 10 nm excitation and 480 ± 10 nm emission wavelengths, with an instrument gain of 1,100. Three independent experiments were performed with separate recombinant protein preparations.

Electron cryo-microscopy. Extracted, pronase-treated tau filaments (3 μl at a concentration of approximately 0.5 mg ml^{-1}) were applied to glow-discharged holey carbon grids (Quantifoil Au R1.2/1.3, 300 mesh) and plunge-frozen in liquid ethane using an FEI Vitrobot Mark IV. Images were acquired on a Gatan K2-Summit detector in counting mode using an FEI Titan Krios at 300 kV. A GIF-quantum energy filter (Gatan) was used with a slit width of 20 eV to remove inelastically scattered electrons. In total, 52 movie frames were recorded, each with an exposure time of 250 ms using a dose rate of 1.06 electrons per Å^2 per frame, for a total accumulated dose of 55 electrons per Å^2 at a pixel size of 1.15 Å on the specimen. Defocus values ranged from −1.7 to −2.8 μm. Further details are presented in Extended Data Table 3.

Helical reconstruction. Movie frames were corrected for gain reference, motion-corrected and dose-weighted using MOTIONCOR2[38]. Aligned, non-dose-weighted micrographs were used to estimate the contrast transfer function (CTF) in Gctf[39]. All subsequent image-processing steps were performed using helical reconstruction methods in RELION 2.1[22,40]. NPFs and WPFs were picked manually and processed as separate datasets.

NPF dataset. NPF segments were extracted using a box size of 270 pixels and an inter-box distance of 14 Å. Reference-free 2D classification was performed using a regularization value of $T=2$, and segments contributing to suboptimal 2D class averages were discarded. An initial helical twist of −0.73° was estimated from the crossover distances of NPFs in micrographs, and the helical rise was estimated to be 4.7 Å. Using these values, an initial 3D reference was reconstructed de novo from 2D class averages of segments comprising an entire helical crossover. A first round of 3D classification, starting from the de novo initial model low-pass filtered to 40 Å, with local optimization of the helical twist and rise, and a regularization value of $T=4$ yielded a reconstruction in which individual β-sheets perpendicular to the helical axis were clearly separated, but no structure was discernable along the helical axis. Subsequently, 3D auto-refinement with optimization of helical twist and rise and a regularization value of $T=10$ was performed using the segments that contributed to the 3D class displaying β-sheets. The resulting reconstruction showed clearly discernible β-strand separation.

An additional round of 3D classification with a regularization value of $T=10$ starting from the 5 Å low-pass filtered map from the previous auto-refinement was used to select further segments for a final high-resolution refinement. In total, 16,097 segments contributed to the final map. The reconstruction obtained with this relatively small subset of the initial dataset matched lower-resolution reconstructions obtained with larger subsets of the data, indicating that image classification did not select for a specific structure from a conformationally heterogeneous dataset, but instead was successful in distinguishing the segments with high-resolution information from images of varying quality. This is in line with observations in single-particle analysis[41]. Superimposing the selected segments onto the original micrographs further confirmed this. Image classification also did not separate filaments with variable twists; instead, RELION combines segments from filaments with variable twists into a single 3D reconstruction and reduces the corresponding blurring effects by only using the central part of an intermediate asymmetrical reconstruction for real-space helical symmetrisation[22]. We used a 10% value for the corresponding helical z_percentage parameter

Optimization of the helical twist and rise converged onto −0.75° and 4.78 Å, respectively. Refinements with helical rises of multiples of 4.78 Å all led to β-strand separation, but in agreement with the observed absence of layer lines between 50 and 4.7 Å we were unable to detect any repeating patterns along the helical axis other than the successive rungs of β-strands.

The final NPF reconstruction was sharpened using standard post-processing procedures in RELION, resulting in a *B*-factor of −57 Å^2 (Extended Data Table 3). Helical symmetry was imposed on the post-processed map using RELION helix toolbox[22]. Final, overall resolution estimates were calculated from Fourier shell correlations at 0.143 between the two independently refined half-maps, using phase-randomization to correct for convolution effects of a generous, soft-edged solvent mask[42]. The overall resolution estimate of the final map was 3.2 Å. Local resolution estimates were obtained using the same phase-randomization procedure, but with a soft spherical mask that was moved over the entire map.

WPF dataset. The WPF dataset was down-scaled to a pixel size of 3.45 Å and segments were extracted using a box size of 180 pixels and an inter-box distance of 14 Å. As with the NPF dataset, an initial 3D reference was reconstructed de novo from 2D class averages of segments comprising an entire helical crossover. 3D classification was then performed to discard suboptimal segments. 3D auto-refinement of the best class with a regularization value of $T=4$ and a fixed helical rise and twist of 4.7 Å and −0.6°, respectively, led to a 3D structure with good separation of β-sheets perpendicular to the helical axis, but no structure was discernible along the helical axis. The cross-section of this map clearly revealed the presence of two NPF protofilaments. To further improve the reconstruction, we also made an initial model by placing two NPF maps, rotated 180° relative to each other in the WPF reconstruction, and low-pass filtering the resulting map to 60 Å. After a second 3D auto-refinement starting from this model, the final WPF reconstruction had an estimated overall resolution of 8 Å and was sharpened by specifying a *B*-factor of −200 Å^2 (Extended Data Table 3). In total, 3,003 segments contributed to the final map.

Model building and refinement. A single monomer of the NPF core was built de novo in the 3.2 Å resolution reconstruction using COOT[43]. Model building was started from the distinctive extended β-spiral conformation of the 332-Pro-Gly-Gly-Gly-335 motif, neighbouring the large histidine side chains of residues 329 and 330, and working towards the N- and C-terminal regions by manually adding amino acids, followed by targeted real-space refinement. This assignment of amino acids was unambiguous due to the clear densities of the three Pro-Gly-Gly-Gly motifs and the large aromatic side chains of Tyr310 and Phe346. The model was then translated to give a stack of three consecutive monomers to preserve nearest-neighbour interactions for the middle chain in subsequent refinements using a combination of real-space refinement in PHENIX[44] and Fourier-space

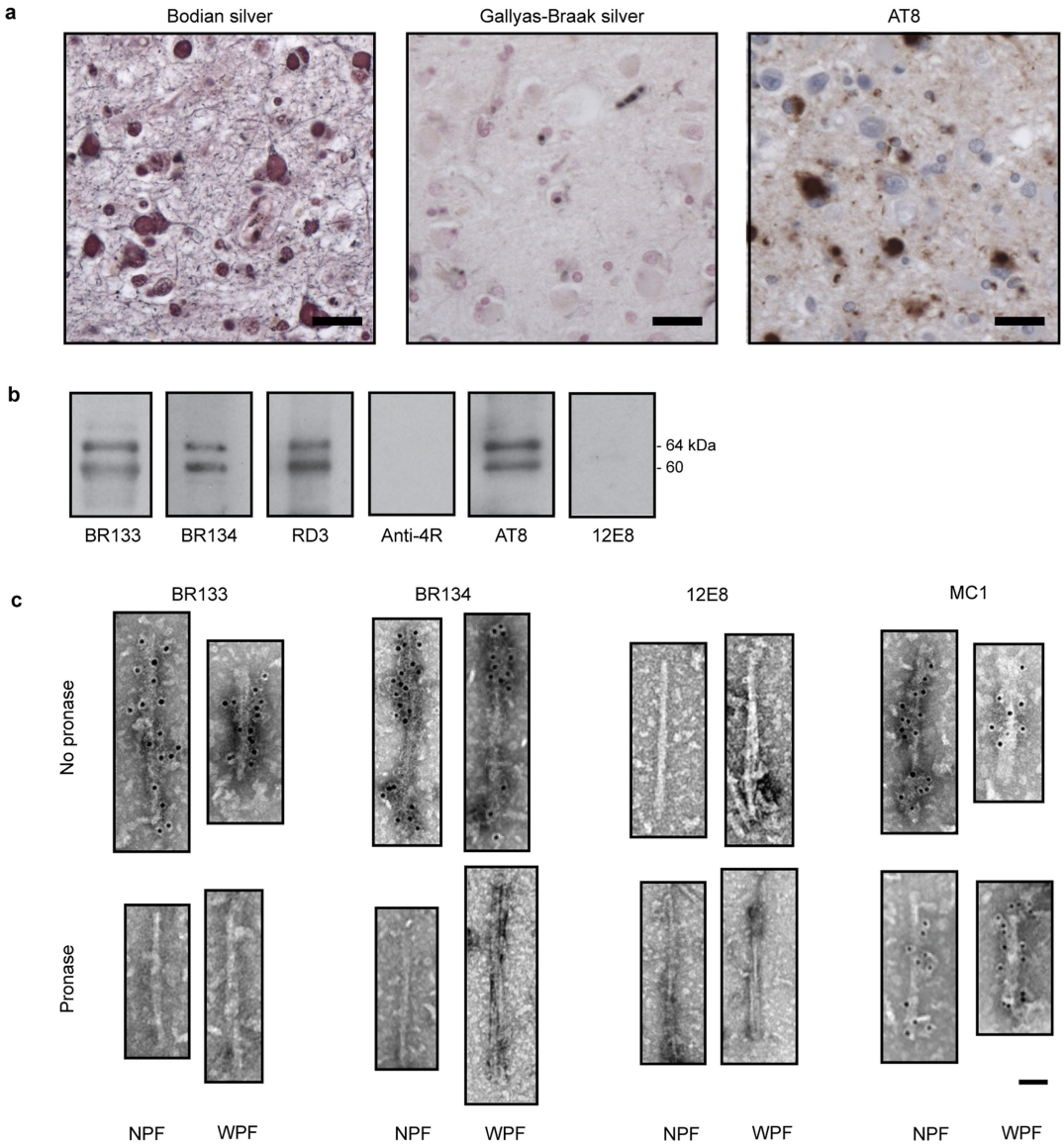
refinement in REFMAC[45]. In the latter, local symmetry restraints were imposed to keep all β-strand rungs identical. Because most of the structure adopts a β-strand conformation, hydrogen bond restraints were imposed to preserve a parallel, in-register hydrogen bonding pattern in earlier stages of the model building process. Side-chain clashes were detected using MOLPROBITY[46] and corrected by iterative cycles of real-space refinement in COOT and Fourier-space refinement in REFMAC. The refined model of the NPF was rigid-body fitted into the WPF map. Separate NPF model refinements were performed against a single half-map, and the resulting model was compared to the other half-map to confirm the absence of overfitting. The final model was stable in refinements without additional restraints.

Ethical review board and informed consent. The Indiana Alzheimer Disease Center studies were reviewed and approved by the Indiana University Institutional Review Board. Informed consent was obtained from the patients' next of kin.

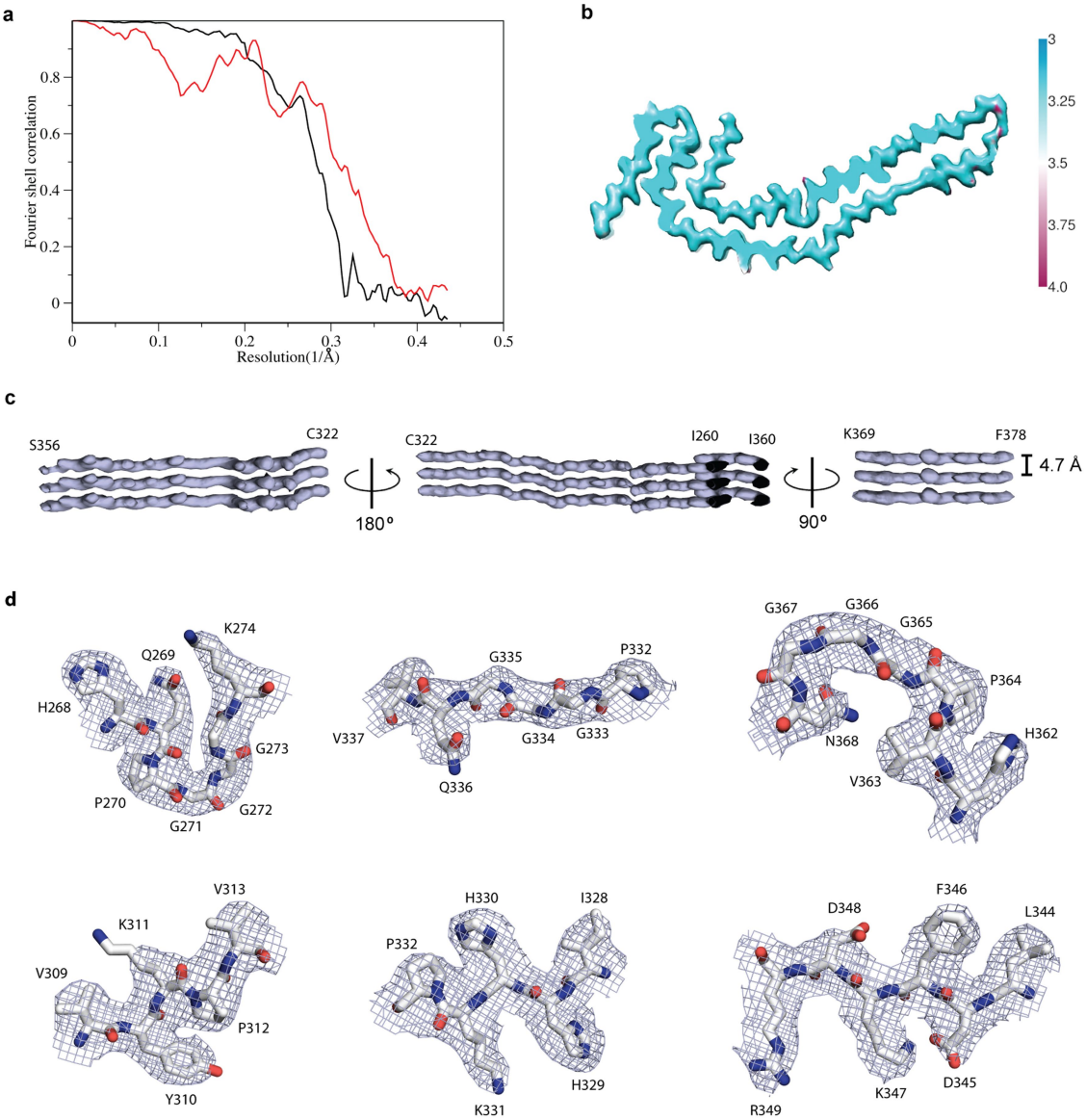
Reporting summary. Further information on experimental design is available in the Nature Research Reporting Summary linked to this paper.

Data availability. Cryo-EM maps have been deposited in the Electron Microscopy Data Bank (EMDB) under accession numbers EMD-0077 for NPF and EMD-0078 for WPF. The NPF refined atomic model has been deposited in the Protein Data Bank (PDB) under accession number 6GX5. Whole-exome sequencing data that support the findings of this study have been deposited in the European Genome-Phenome Archive (EGA), which is hosted by the EBI and the CRG, under accession number EGAS00001003106.

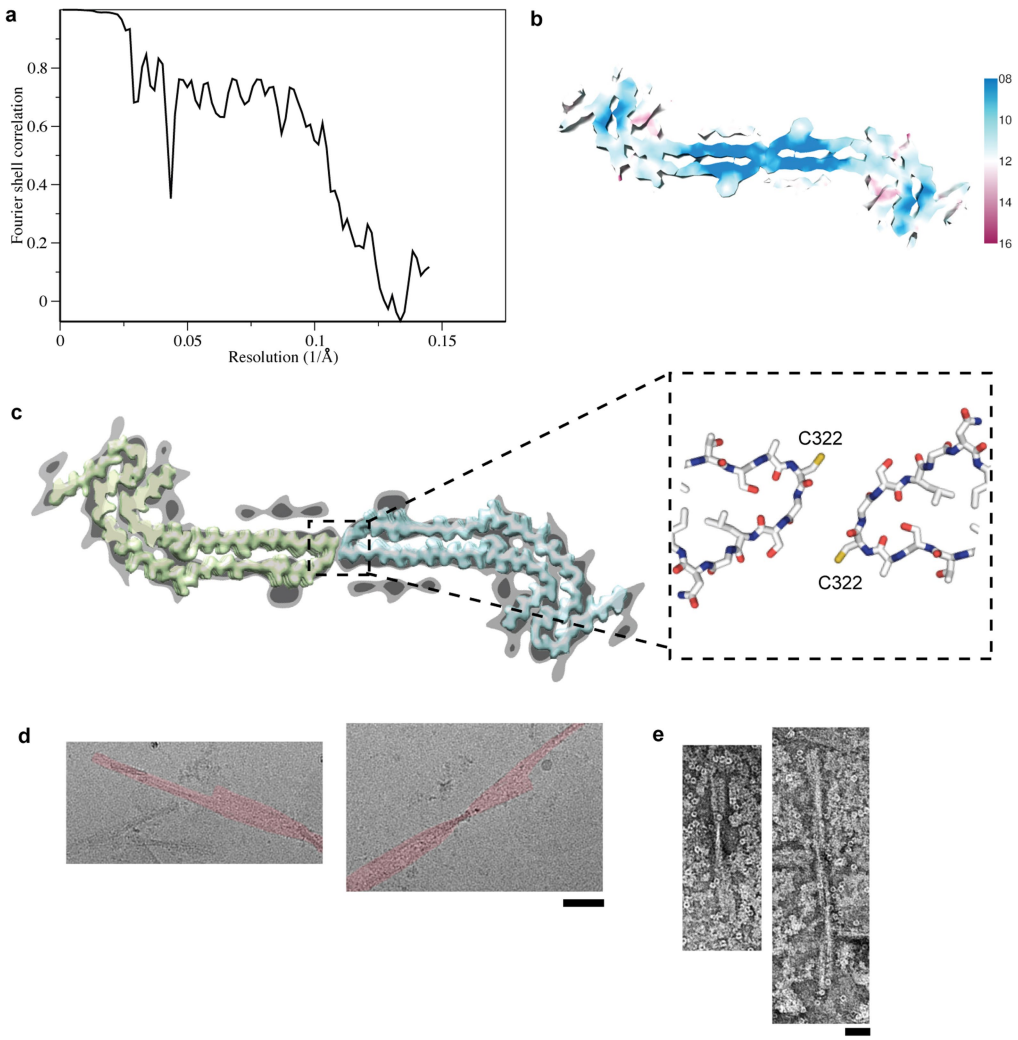
33. Goedert, M., Spillantini, M. G., Cairns, N. J. & Crowther, R. A. Tau proteins of Alzheimer paired helical filaments: abnormal phosphorylation of all six brain isoforms. *Neuron* **8**, 159–168 (1992).
34. Falcon, B. et al. Conformation determines the seeding potencies of native and recombinant Tau aggregates. *J. Biol. Chem.* **290**, 1049–1065 (2015).
35. Spina, S. et al. The tauopathy associated with mutation +3 in intron 10 of Tau: characterization of the MSTD family. *Brain* **131**, 72–89 (2008).
36. Farlow, J. L. et al. Whole-exome sequencing in familial Parkinson disease. *JAMA Neurol.* **73**, 68–75 (2016).
37. Saijo, E. et al. Ultrasensitive and selective detection of 3-repeat tau seeding activity in Pick disease brain and cerebrospinal fluid. *Acta Neuropathol.* **133**, 751–765 (2017).
38. Zheng, S. Q. et al. MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy. *Nat. Methods* **14**, 331–332 (2017).
39. Zhang, K. Gctf: Real-time CTF determination and correction. *J. Struct. Biol.* **193**, 1–12 (2016).
40. Scheres, S. H. RELION: implementation of a Bayesian approach to cryo-EM structure determination. *J. Struct. Biol.* **180**, 519–530 (2012).
41. Fernandez-Leiro, R. & Scheres, S. H. Unravelling biological macromolecules with cryo-electron microscopy. *Nature* **537**, 339–346 (2016).
42. Chen, S. et al. High-resolution noise substitution to measure overfitting and validate resolution in 3D structure determination by single particle electron cryomicroscopy. *Ultramicroscopy* **135**, 24–35 (2013).
43. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. D* **66**, 486–501 (2010).
44. Adams, P. D. et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D* **66**, 213–221 (2010).
45. Murshudov, G. N., Vagin, A. A. & Dodson, E. J. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. D* **53**, 240–255 (1997).
46. Chen, V. B. et al. MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. D* **66**, 12–21 (2010).
47. Seubert, P. et al. Detection of phosphorylated Ser262 in fetal tau, adult tau, and paired helical filament tau. *J. Biol. Chem.* **270**, 18917–18922 (1995).
48. Litersky, J. M. et al. Tau protein is phosphorylated by cyclic AMP-dependent protein kinase and calcium/calmodulin-dependent protein kinase II within its microtubule-binding domains at Ser-262 and Ser-356. *Biochem. J.* **316**, 655–660 (1996).



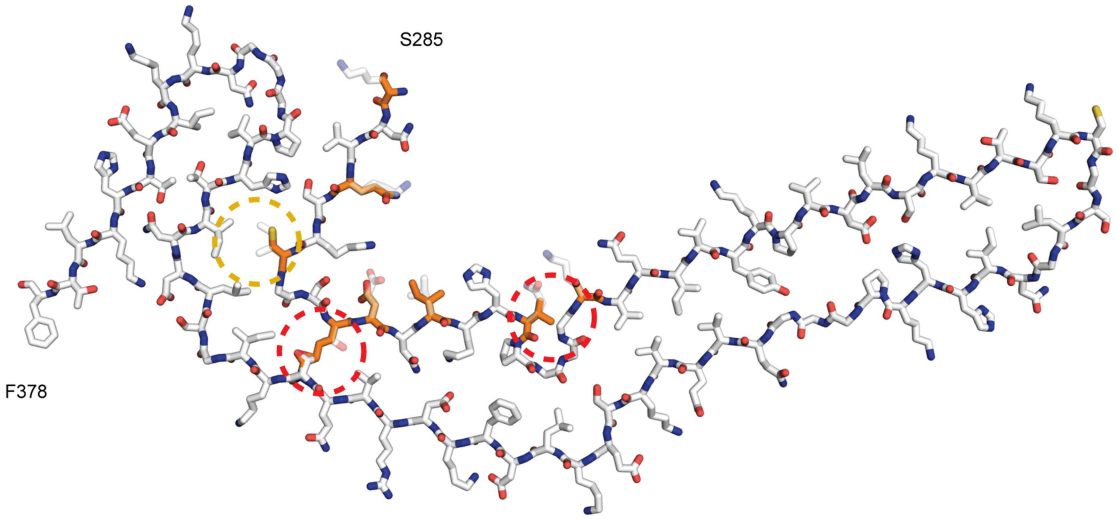
Extended Data Fig. 1 | Further characterization of the filamentous tau pathology of Pick's disease. a, Staining of Pick bodies in the frontotemporal cortex of patient 4 by Bodian silver and antibody AT8 (pS202 and pT205 tau), but not by Gallyas–Braak silver. Nuclei are counterstained blue in the AT8 panel. Scale bars, 20 μm. **b**, **c**, Immunolabelling of the tau filaments extracted from the frontotemporal cortex of patient 4. **b**, Immunoblots with antibodies BR133 (tau N terminus), BR134 (tau C terminus), RD3 (3R tau), anti-4R (4R tau), AT8 (pS202 and pT205 tau) and 12E8 (pS262 tau and/or pS356 tau). **c**, Immunogold negative-stain electron micrographs with antibodies BR133, BR134, 12E8 and MC1 of NPFs and WPFs with and without pronase treatment. Scale bar, 500 Å.



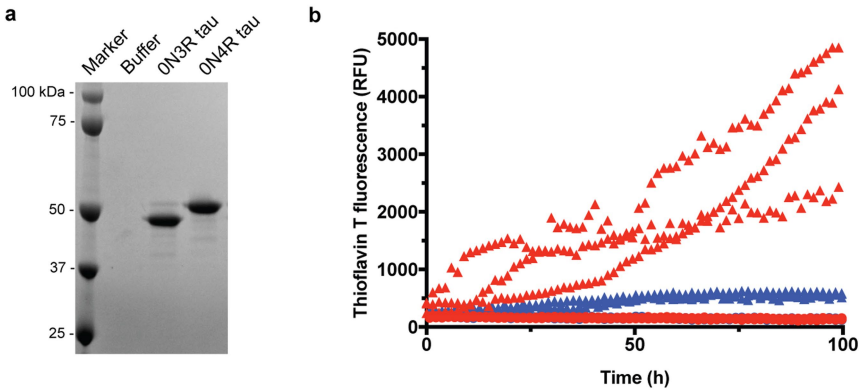
Extended Data Fig. 2 | NPF structure. a, Fourier shell correlation curves between two independently refined half-maps (black line) and between the cryo-EM reconstruction and refined atomic model (red line). **b**, Local resolution estimates for the NPF reconstruction. **c**, Helical axis views of the NPF reconstruction. **d**, Close-up views of the cryo-EM map with the atomic model overlaid. The top row shows the three Pro-Gly-Gly-Gly (PGGG) motifs; the bottom row shows several amino acids with large side chains.



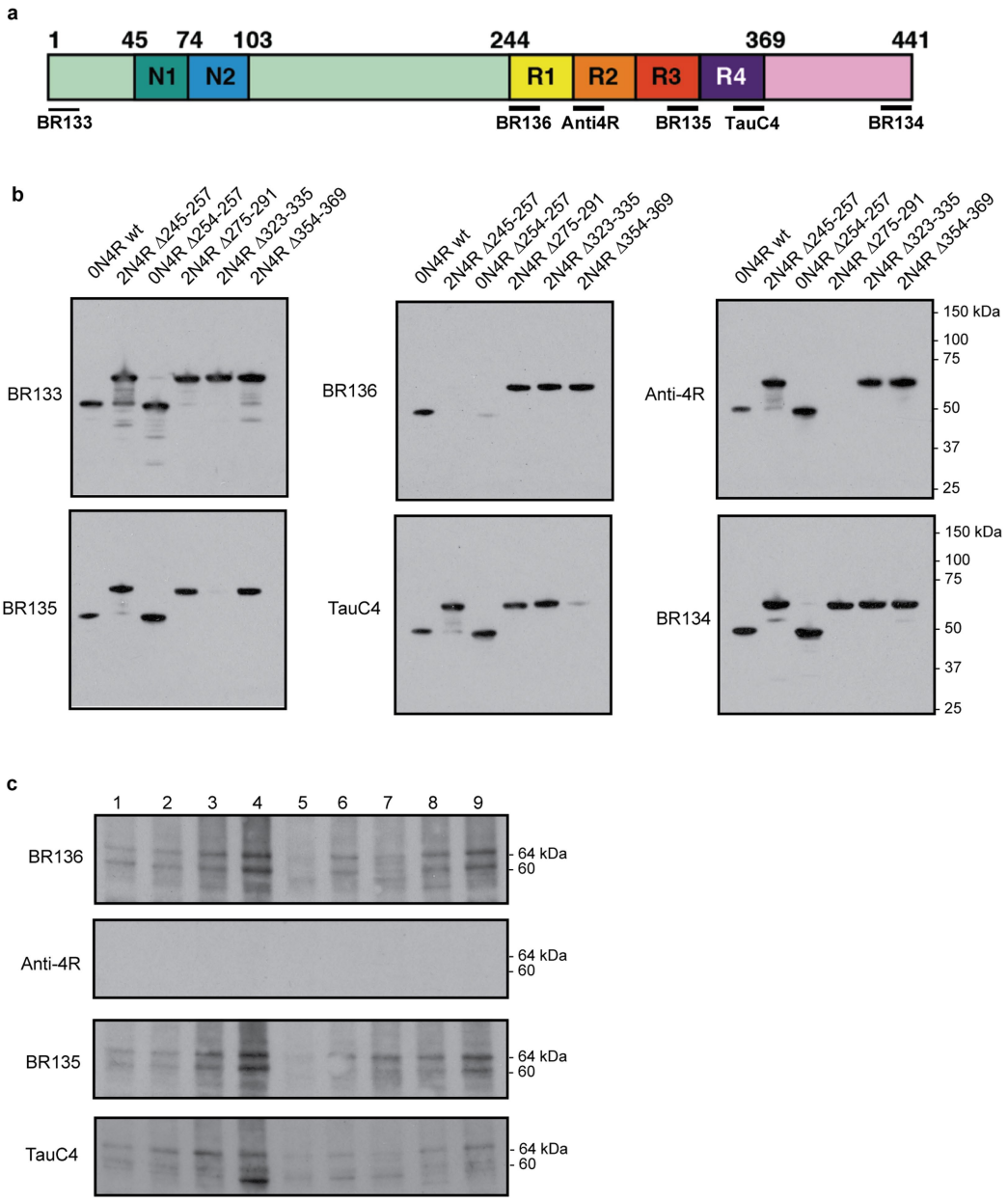
Extended Data Fig. 3 | WPF structure. a, Fourier shell correlation curves between two independently refined half-maps. **b**, Local resolution estimates for the WPF reconstruction. **c**, WPF density at high (light grey) and low (dark grey) threshold with densities for two NPFs overlaid (yellow and blue). The atomic models fitted to the NPF densities in the region of the protofilament interface are shown in the boxed out area. **d**, Cryo-EM images showing WPFs from patient 4 (false coloured red), in which segments from one of the protofilaments have been lost. Scale bar, 500 Å. **e**, Negative-stain EM images of WPFs from patient 4 after incubation in 100 mM dithiothreitol for 20 h. Scale bar, 500 Å.



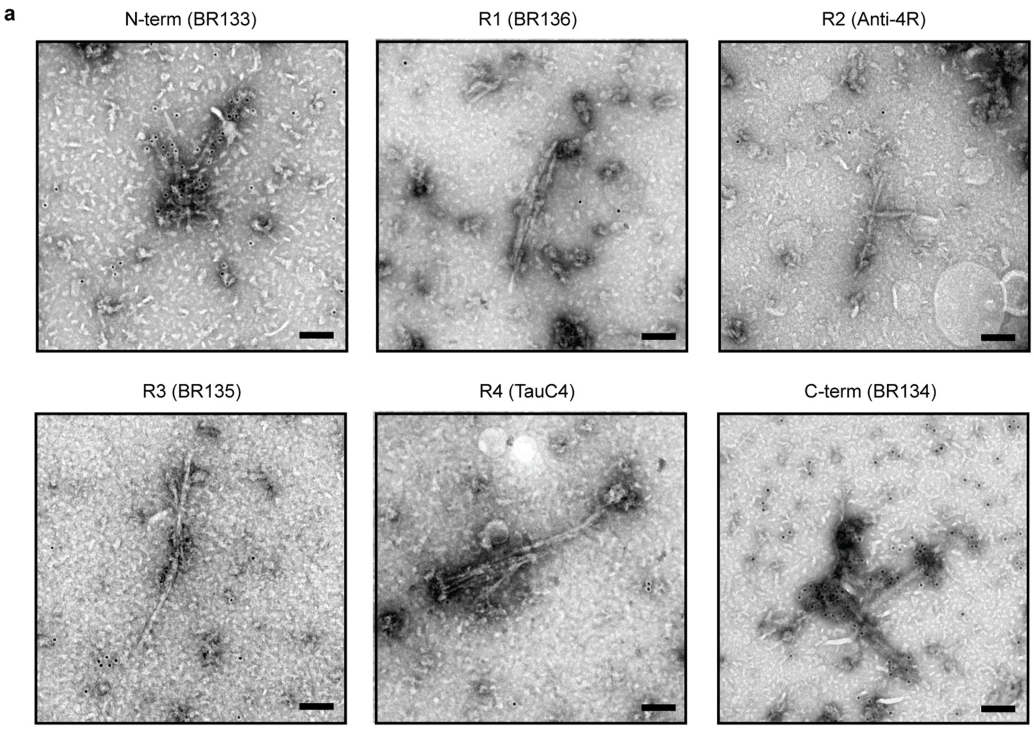
Extended Data Fig. 4 | Incompatibility of the Pick tau filament fold with 4R tau. Atomic model of the Pick fold with the 4R tau sequence overlaid. The region formed by Lys254–Lys274 from R1 is replaced by the Ser285–Val300 region from R2 in 4R tau. Residues that differ between these regions of R1 and R2 are coloured orange. The major discrepancies of lysine at position 294 in R2, instead of threonine at position 263 in R1, and valine at position 300 in R2, instead of glutamine at position 269 in R1, are highlighted with dashed red outlines. The minor discrepancy of weaker interactions of Cys291 of R2 with Leu357 and Ile360 compared with those formed by Ile260 of R1 is highlighted with a dashed yellow outline.



Extended Data Fig. 5 | Seeded aggregation of full-length 3R, but not 4R, tau by the sarkosyl-insoluble fraction from the brain of patient 4 with Pick's disease. a, Coomassie-stained SDS–PAGE of the 0N3R and 0N4R recombinant tau preparations used for seeded aggregation. Two additional recombinant tau preparations were performed with similar results. **b**, Thioflavin T fluorescence measurements of 0N3R (red) and 0N4R (blue) recombinant tau after incubation with (triangles) or without (circles) the sarkosyl-insoluble fraction from the frontotemporal cortex of patient 4. The results are from three independent experiments using separate recombinant protein preparations. The sarkosyl-insoluble fraction from Pick's disease brain efficiently seeded the aggregation of 3R, but not 4R, tau. RFU, relative fluorescent units.



Extended Data Fig. 6 | Immunoblot analysis of additional Pick's disease cases. a, Diagram of 2N4R tau showing the N-terminal inserts (N1 and N2), the repeats (R1–R4) and the epitopes of antibodies BR133 (N terminus), BR136 (R1), anti-4R (R2), BR135 (R3), TauC4 (R4) and BR134 (C terminus). **b**, Immunoblots of epitope-deletion recombinant tau constructs with the antibodies shown in **a**. Identical results were obtained in two independent repeats. **c**, Immunoblots using the antibodies BR136, anti-4R, BR135 and TauC4 of tau filaments extracted from the frontotemporal cortex of 9 patients with Pick's disease (patient 4 was used for cryo-EM). See Extended Data Table 1 for details of the patients with Pick's disease.



b

Pick's disease case number	N-term	R1	R2	R3	R4	C-term
	Antibody epitopes					
1	✓	×	×	×	×	✓
2	✓	×	×	×	×	✓
3	✓	×	×	×	×	✓
4	✓	×	×	×	×	✓
5	✓	×	×	×	×	✓
6	✓	×	×	×	×	✓
7	✓	×	×	×	×	✓
8	✓	×	×	×	×	✓
9	✓	×	×	×	×	✓

Extended Data Fig. 7 | Immunogold negative-stain EM analysis of additional Pick's disease cases. a, Representative immunogold negative-stain electron microscopy with antibodies BR133 (tau N terminus), BR136 (tau R1), anti-4R (tau R2), BR135 (tau R3), TauC4 (tau R4) and BR134 (tau C terminus) of NPFs and WPFs extracted from the frontotemporal cortex of patient 4. Scale bars, 100 nm. Similar results were obtained with tau filaments extracted from the frontotemporal cortices of eight additional patients with Pick's disease. **b**, Table summarizing results from **a**. Results for patient 4 (shown in **a** and used for cryo-EM) are highlighted in yellow. See Extended Data Table 1 for details of the patients with Pick's disease. Tick marks indicate antibody decoration of filaments, and crosses indicate that the antibodies did not decorate filaments. NPFs and WPFs were decorated by the antibodies against the N and C termini of tau, but not by the tau repeat-specific antibodies.

Extended Data Table 1 | Summary of the patients with Pick's disease

Patient	Gender	Age at death (years)	*MAPT*	*PSEN1*	*APOE* haplotypes	Post-mortem interval (h)
1	F	73	wt	wt	ε3/ε3	23.8
2	F	70	wt	wt	ε3/ε3	20.5
3	M	61	wt	wt	ε3/ε4	24.5
4	F	63	wt	wt	ε2/ε3	3.0
5	M	70	wt	wt	ε3/ε3	14.0
6	M	65	wt	wt	ε3/ε3	12.5
7	M	64	wt	wt	ε3/ε4	3.0
8	M	56	wt	wt	ε3/ε4	9.0
9	M	69	wt	wt	ε3/ε4	4.5

Wild-type (wt) means that no known disease-causing mutations in the tau gene (*MAPT*) or the presenilin-1 gene (*PSEN1*) were detected. The patient used for cryo-EM is highlighted in yellow.

Extended Data Table 2 | Anti-tau antibodies

Name	Epitope	Supplier	Cat. number	Species	Type	WB dilution	EM dilution	IHC dilution	Validation
BR133	N-terminus	In house	-	Rabbit	Polyclonal	1:4,000	1:50	-	[30]
BR134	C-terminus	In house	-	Rabbit	Polyclonal	1:4,000	1:50	-	[30]
BR136	R1	In house	-	Rabbit	Polyclonal	1:4,000	1:50	-	Extended Data Fig. 6
Anti-4R	R2	Cosmo Bio	CAC-TIP-4RT-P01	Rabbit	Polyclonal	1:2,000	1:50	1:100	Manufacturer's datasheet and Extended Data Fig. 6
BR135	R3	In house	-	Rabbit	Polyclonal	1:4,000	1:50	-	[30] and Extended Data Fig. 6
Tau C4	R4	Masato Hasegawa	-	Rabbit	Polyclonal	1:2,000	1:50	-	[7] and Extended Data Fig. 6
RD3	R1/3	Millipore	05-803	Mouse	Monoclonal	1:4,000	-	1:3,000	Manufacturer's datasheet
12E8	pS262 and/or pS356	Peter Seubert	-	Mouse	Monoclonal	1:100,000	1:50	1:1,000	[47,48]
AT8	pS202 and pT205	Thermo	MN1020	Mouse	Monoclonal	1:1,000	1:50	1:300	Manufacturer's datasheet
MC1	Discontinuous epitope (residues 7-9 and 313-322)	Peter Davies	-	Mouse	Monoclonal	-	1:10	-	[24]

IHC, immunohistochemistry; EM, immunogold negative-stain electron microscopy; WB, western blot.

Extended Data Table 3 | Cryo-EM data collection, refinement and validation statistics

	NPF (EMDB-0077) (PDB 6GX5)	WPF (EMDB-0078)
Data collection and processing		
Magnification	x105,000	x105,000
Voltage (kV)	300	300
Electron exposure (e–/Å^2)	55	55
Defocus range (μm)	-1.7 to -2.8	-1.7 to -2.8
Pixel size (Å)	1.15	3.45
Symmetry imposed	None	C2
Initial particle images (no.)	83,475	8,024
Final particle images (no.)	16,097	3,003
Map resolution (Å)	3.2	8
FSC threshold	0.143	0.143
Map resolution range (Å)	3.2 to 3.5	8 to 12
Helical rise (Å)	4.78	4.7
Helical twist (°)	-0.75	-0.6
Refinement		
Initial model used (PDB code)	*de novo*	n/a
Model resolution (Å)	3.2	n/a
FSC threshold	0.5	
Model resolution range (Å)	3.2	n/a
Map sharpening *B* factor (Å^2)	-57	-200
Model composition		n/a
Non-hydrogen atoms	2133	
Protein residues	282	
Ligands	0	
B factors (Å^2)		n/a
Protein	75.8 average	
Ligand	n/a	
R.m.s. deviations		n/a
Bond lengths (Å)	0.0175	
Bond angles (°)	1.1678	
Validation		n/a
MolProbity score	2.15	
Clashscore	5.96	
Poor rotamers (%)	4.94	
Ramachandran plot		n/a
Favored (%)	95.65	
Allowed (%)	100	
Disallowed (%)	0	

ROCKET LAB

Space firm Rocket Lab, whose launch site on New Zealand's Mahia peninsula is shown here, typifies the country's growing expertise in space technology.

MOBILITY

A land poised for lift-off

New Zealand's embrace of science promises to make it a rewarding career destination.

BY JAMES MITCHELL CROW

At high tide, the Cook Strait — the ribbon of water separating New Zealand's North and South Islands — is still and calm. Aboard the research vessel *Tangaroa*, spatial ecologist Kim Goetz seizes the moment before the tide turns, and sends an acoustic recorder plunging downwards. The recorder will spend the next six months moored to the seabed, capturing the calls of nearby marine mammals.

Coordinating New Zealand's first large-scale deployment of acoustic recorders across the strait was one of Goetz's first tasks after joining New Zealand's National Institute of Water and Atmospheric Research (NIWA) in Wellington. "More than half the world's species of whales and dolphins come through New Zealand waters, and there's next to nothing known about them," says Goetz, who left her native United States in 2014 to join NIWA. A canyon along the strait's sea floor potentially makes it a popular hangout for deep-diving species.

Using the data collected from the recorders, Goetz has confirmed that the region is a hotspot for rare mammals, including elusive beaked whales (belonging to the family Ziphiidae). She says that the whales' deep-sea habitat had made it difficult for researchers to learn much about them except when they wash ashore — until now. The team has also deployed acoustic moorings farther north, to monitor critically endangered Maui dolphins.

NIWA and other New Zealand research organizations have seen sustained funding growth in the past decade, as the country has increasingly focused on supporting basic and applied science. In 2008, the government appointed its first chief scientific adviser, a role that University of Auckland biochemist Juliet Gerrard assumed on 1 July this year. Gerrard says that she hopes to encourage ambition in young female scientists. The nation's federal science expenditure has increased by more than 70% over the past decade to NZ$1.58 billion (roughly US$1 billion) and the number of full-time equivalent research ▶

positions has grown by about 23% since 2010, to around 25,000.

A new national government, elected in late 2017 and headed by progressive Prime Minister Jacinda Ardern, has yet more-ambitious plans. Among its policies is making New Zealand a net-zero carbon emitter by 2050. Existing technology won't achieve that aim, says Peter Crabtree, general manager for science and innovation at the country's Ministry of Business, Innovation and Employment. One specific target, which could further expand researchers' employment opportunities, is to boost research-and-development investment from 1.28% of gross domestic product to 2% within a decade, bringing it closer to the 2.38% average for member nations of the Organisation for Economic Co-operation and Development.

"There's a strong focus on increasing the proportion of investment from business," Crabtree says. In the new government's first annual budget, NZ$1 billion was set aside over the next four years for an R&D tax incentive. The strategy to boost business R&D also emphasizes the need to attract innovative international businesses into the country — an approach that is already yielding results, including commercial-satellite launches.

The nation's distance from populous research hubs, its high cost of living and a scientific workforce that is only now beginning to grow might prove insurmountable obstacles for some, however.

ROOM TO EXPLORE

The nation's research community has a collaborative, can-do attitude — as Goetz's Cook Strait survey attests. "I'm not an acoustician — in the United States, nobody would have asked me to do that kind of thing," says Goetz. But in New Zealand, whose total population is less than that of Sydney, Australia, there is a smaller pool of researchers to draw from. "You can dabble in a lot of different areas," she says. "You are always learning."

"There's fewer of us around, so we have to cover more bases," agrees Charles Eason, chief executive of the Cawthron Institute in Nelson, the country's largest independent science organization. The Cawthron aims to support the country's economy through science while preserving the natural environment — in which New Zealand's powerful indigenous Maori traditions are deeply rooted. "Our Maori culture plays through our psyche," Eason says. "Maori culture is very strong in terms of environmental protection."

Nearly 15% of New Zealand's population identify as Maori, yet Maori hold only 5% of academic research positions, says Anne-Marie Jackson, an indigenous New Zealander who researches Maori physical education, health and world view at the University of Otago. Still, she is seeing aspects of Maori culture seep into research culture. "The way we think about the environment has shifted," Jackson says. Rather than considering a complex ecosystem in small pieces, researchers now try to take a whole-system approach. "That aligns with one of our core principles of *kaitiakitanga*, a holistic way of looking at anything," Jackson says.

LESS IN THE POT

One downside of a small workforce and employer base is that fewer grants are available than in a larger nation. Nick Golledge, an ice-sheet modeller and climate researcher at the Antarctic Research Centre at Victoria University of Wellington, says that he often loses talented graduate students as they come closer to earning their PhD, because he can't find postdoctoral funding for them. "It's a small country, and the overall amount of money is small," says Golledge.

Others warn that research positions in highly specialized disciplines might not be available. "It is not easy to plan to come here," says Olaf Morgenstern, a meteorologist at NIWA. "If you have a speciality that is not widespread, you might struggle to find a position." Morgenstern had no plans to live in New Zealand, but he decided to move there from the United Kingdom when the right job happened to come up.

Goetz arrived in New Zealand after a similar decision, and says that only in hindsight did she realize how fortunate she was to land a permanent position. "I took it for granted coming out here," she says. "Since then, I've had several friends that had a job on a year term, and they have to leave because they can't find another job. I think that's more common than not."

The country's remote location can make itself felt in the cost of living. And property prices have risen sharply in the past decade — especially in Auckland, a global property market hotspot. Wellington, Christchurch and Dunedin were also ranked "severely unaffordable" in a 2018 international survey, meaning that houses there cost, on average, more than five times the median household income.

Still, the diversified research landscape may beckon for some. New Zealand is also a geological hotspot, thanks largely to the fault line between the Pacific and Australian plates that runs up the country's spine. "Earthquake research, the geomorphology of the landscape — New Zealand is a wonderful laboratory for that kind of work," says James Metson, deputy vice-chancellor for research at the University of Auckland.

SPACE RACE

The government's 2015 analysis of the nation's science strengths identified engineering, computer science, energy research, and physics and astronomy as particular specialisms — and ones that could offer a strong return from increased investment. Dividends are already paying off in one of those areas. In 2016, just five months after setting up a national space agency, New Zealand's parliament created a regulatory framework for space launches. And in January 2018, the US–New Zealand company Rocket Lab launched its first commercial payloads into orbit.

Although most of its funding comes from US venture capitalists, Rocket Lab was founded in Auckland in 2006 by New Zealand scientist and engineer Peter Beck. Its signature Electron rocket, developed and built in Auckland, has been designed to capitalize on the soaring demand for sending small satellites into space.

Scientists who have been in New Zealand for some time have few complaints. "A lot of us coming in realize we will probably never own a house," says Goetz. But there are advantages to the New Zealand lifestyle, she says. "People here really have a value for personal time, and for family time," she says. "And there's a lot of open spaces, beautiful scenery and things to get out and do."

"You could perceive our distance as a disadvantage," says Eason. "But people are pretty keen to come here." ■

James Mitchell Crow *is a freelance science writer based in Melbourne, Australia.*

GOOD TO GO

What to consider if you're moving to New Zealand

Scientists born overseas need a visa to live and work in New Zealand (see go.nature.com/nz_visas for details). Here's a summary of the main categories available.

- Some science-related jobs, including postdoctoral researcher and university lecturer, are on New Zealand's Immediate Skill Shortage list. Those offered a job on this list can apply for the Essential Skills work visa, which grants temporary residency (see go.nature.com/nz_essential).
- Some science-related jobs, including environmental research scientist and food technologist, are on the Long Term Skill Shortage list. Scientists who are granted a Long Term Skill Shortage list work visa can apply for permanent residency (see go.nature.com/nz_longterm).
- New Zealand also has a skilled-migrant visa category, which grants permanent residency (see go.nature.com/nz_migrant). This visa class is points-based, with applicants accruing points on the basis of criteria such as age, skills, qualifications and experience. Applicants with qualifications on the Long Term Skill Shortage list, and applicants who have worked in New Zealand on a temporary visa, gain extra points. J.M.C.

FAILSAFES

Be prepared.

BY STEWART C. BAKER

ILLUSTRATION BY JACEY

1. PLAN FOR THE LONG TERM

Once, the ruin was a city called Toronto, and its buildings spread horizon to horizon, a safe habitat for people like Jeydah.

Now it is only the ruin: a maze of collapsed masonry and broken machines. Jeydah visits once a month to gather salvage and pick berries from the sprawling brambles that grow in its once-proud avenues. The people in the plains villages think her strange, but they barter with her all the same, and she stays fed on the tools that she finds, the gadgets she makes.

The best salvage, though, Jeydah keeps for herself: a sphere that gives off light; a pane of glass you can heat with a touch; a small black block that speaks stories of the way the world once was. She keeps these treasures in a weatherproof sack, safely hidden in a hollow tree half-a-day's walk from the ruin.

Or perhaps not so safely, because when she stops at her tree one chill autumn day, the sack is missing, her treasures gone. The breaks in the long stalks of grass nearby look recent, so Jeydah hikes on without stopping. No matter how badly she wants her things back, few people in the wastes are kind to strangers.

It's dark when she reaches the ruin, and for the first time in all her visits, she notices something new. Light — pale and steady as her stolen sphere's — pooled inside a narrow arch at the ruin's outermost edge.

2. ELIMINATE BARRIERS TO ACCESS

The arch leads to a stairwell, which leads to a warren of tunnels. As Jeydah walks, arrows light the floor, pointing her down a path she would never have found without them.

At any other time, she would have marvelled at this ancient tech, but tonight her shoulders ache with tension. Although she has not encountered whoever stole her things, she saw dust rising from the dry road behind her as she fled her tree. She jogs through the tunnels as fast as she dares, hoping her pursuer is more timid than she is.

After countless twists and turns, the arrows dead-end against a wall — solid, smooth and indistinguishable from any other in this place. Jeydah spins around, throat tight, but only manages three steps before the tunnels plunge into darkness.

At least she knows she wasn't followed.

She laughs once, bleakly, then feels her way back to the dead end and leans against it, head bent, eyes closed. When the wall dissolves, she falls right through.

3. BREAK DOWN COMPLEX TASKS

The room behind the wall is large and open, suffused with a pale blue glow that leaves many of its features in shadow. More arrows lead Jeydah to the far wall, which whirrs open to reveal a bowl. The smell of chillies and roasted grain meets her nose, and her stomach growls.

She should be cautious, but she's tired and hasn't eaten in nearly a day. Besides, she'd be dead already if whoever built this place wanted it. She lifts the bowl from its platform.

The arrows change, leading her to a chair lurking in the shadows. The instant she sits, the wall flickers to life with staticky symbols. She can't read them, of course, but there's a picture of a person in the chair pushing a button and that seems clear enough.

She looks down, finds the button, and pushes it.

The lights in the room increase, and the wall displays more symbols, more pictures: the person in the chair retrieves objects from the wall, puts them together to build tool after tool. The person multiplies until a whole community tills the land, builds structures, rides wagons that move themselves.

Jeydah stares, entranced, until the wall flickers and shows a single picture of the archway, a black panel and a lightning bolt that's cut in half.

The wall whirs open, revealing a panel. Jeydah picks it up, looks at the picture on the wall, and heads into the labyrinth.

4. ALWAYS HAVE A BACK-UP

Jeydah takes a few hesitant steps into the tunnels, relaxing only when she sees the arrows appear beneath her feet, pointing her onwards.

After that, it takes no time at all to reach the stairs and climb to the surface, where the Moon hangs low and full above the ruin. Not much longer to clear the roof behind the arch of the debris that clutters it.

Black panels glimmer in the moonlight, one of which is cracked. She pulls it free and replaces it with the new board, then slips back below the surface.

5. MAKE PAST RESULTS REPRODUCIBLE

As she follows the arrows back through the tunnel, Jeydah hears hiccupping sobs, high-pitched and desperate, that seem to come from everywhere at once.

In all her explorations of the ruin, Jeydah's never once met a ghost, so she squares her shoulders and follows the sound to a small, hungry-looking girl, hunched over a sack that looks awfully familiar and crying her eyes out. The child looks up as Jeydah approaches, clutches the sack to her bone-thin body, and scurries back against the tunnel wall, chin held up defiantly.

If Jeydah had found the little thief at her clearing, she would have snatched the sack back in an instant. Fought the girl if she had to. Harsh, but the wastes are harsh. The world is.

After the wonders of the labyrinth, though, after the secret room with its hot food and its worlds-ago tech, that seems petty. Unjust. She remembers the pictures on the wall. Remembers how hungry she was as a child.

She squats down next to the girl, relaxes her shoulders, puts on her most disarming grin. "Hey," she says. "You want something to eat?" ■

Stewart C. Baker *is an academic librarian, haikuist and speculative-fiction writer based in Oregon. His fiction has appeared in* Daily Science Fiction, Flash Fiction Online, Nature *and other magazines.*

NATURE.COM
Follow Futures:
@NatureFutures
go.nature.com/mtoodm